



**HAL**  
open science

# Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices

Nicolas Debarsy, James Lesage

► **To cite this version:**

Nicolas Debarsy, James Lesage. Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices. *Journal of Business and Economic Statistics*, 2022, 40 (2), pp.547-558. 10.1080/07350015.2020.1840993 . hal-03046651

**HAL Id: hal-03046651**

**<https://hal.science/hal-03046651v1>**

Submitted on 11 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices

Nicolas Debarsy  
CNRS- LEM UMR 9221  
Université de Lille  
Campus Cité Scientifique  
Bat. SH2 F-59655 Villeneuve-d'Ascq  
nicolas.debarsy@cnrs.fr  
and  
James P. LeSage  
Texas State University  
Department of Finance & Economics  
601 University Drive  
San Marcos, TX 78666, USA  
james.lesage@txstate.edu

## Abstract

There is a great deal of literature regarding use of non-geographically based connectivity matrices or combinations of geographic and non-geographic structures in spatial econometric models. We focus on convex combinations of weight matrices that result in a single weight matrix reflecting multiple types of connectivity, where coefficients from the convex combination can be used for inference regarding the relative importance of each type of connectivity in the global cross-sectional dependence scheme. We tackle the question of model uncertainty regarding selection of the best convex combination by Bayesian model averaging. We use Metropolis-Hastings guided Monte Carlo integration during MCMC estimation of the models to produce log-marginal likelihoods and associated posterior model probabilities. We focus on MCMC estimation, computation of posterior model probabilities, model averaged estimates of the parameters, scalar summary measures of the non-linear partial derivative impacts, and their associated empirical measures of dispersion.

KEYWORDS: Markov Chain Monte Carlo estimation, SAR, block sampling parameters for a convex combination, cross-sectional dependence, hedonic price model.

# 1 Introduction

Spatial regression models typically rely on spatial proximity to specify connectivity matrices, where the relative Euclidean distance between observations determines the strength of dependence between observations. One can generalize the notion of Euclidean distance to produce measures of dependence between observations based on other metrics. For example, Pace et al. (2000) proposed a model for prices of homes sold that occur at irregular points in space and time, generalizing distance to include relative locations in time. Related work by Pace et al. (2002) relied on generalized distances that considered the number of bedrooms and bathrooms (of nearby homes) to specify the structure of selling price dependence between homes, with the motivation that appraisers determine market price estimates based on homes comparable in these two metrics.

Once we open the door to non-spatial metrics as a way to specify dependence between cross-sectional observations, a host of issues arise, which are discussed in LeSage and Pace (2011) and Debarsy and LeSage (2018). Blankmeyer et al. (2011, p.94) point out that *“a single weight matrix, based on a multivariate similarity criterion (generalized distance) requires a norm to prevent scale differences from influencing the weight placed on the various measures of similarity. (This is unlike the case of spatial proximity where Euclidian distance provides a natural scaling.)”*

We avoid the scaling issue that arises in the case of generalized distance using an approach set forth in Debarsy and LeSage (2018) that relies on convex combinations of different connectivity matrices to form a single weight matrix, first explored by Pace and LeSage (2002) as well as Hazır et al. (2018). The convex combination approach uses a single  $n \times n$  weight matrix that represents the global cross-sectional dependence scheme. This combination is constructed from alternative underlying types of connectivity between  $n$  observations reflected by  $n \times n$  matrices which capture different modeling strategies for cross-sectional dependence. The parameters associated with each matrix in the convex

combination indicate the relative importance assigned to each type of dependence in the global cross-sectional dependence scheme. This approach avoids the issue of scaling for different metrics used in a generalized measure of distance as each connectivity matrix entering the convex combination is normalized beforehand.

In this contribution, we focus on the spatial autoregressive model with a convex combination of connectivity matrices and follow Debarsy and LeSage (2018) in using Markov Chain Monte Carlo (MCMC) estimation of the model. However, in contrast to Debarsy and LeSage (2018), this paper explicitly addresses the estimation, inference and numerical challenges raised by SAR models with convex combination of connectivity matrices. Debarsy and LeSage (2018) discuss issues pertaining to estimation and interpretation of models involving convex combinations of only two connectivity matrices, which do not require the approach developed here. We tackle the problem of model uncertainty inherent in models involving numerous connectivity matrices by relying on a Bayesian model averaging procedure. Further, the sampling procedure for the convex combination parameters requires specific treatment since they have to be jointly drawn and need to obey two sets of constraints. Also, we develop measures of dispersion for the effects of interest (direct; indirect and total) that account for uncertainty in estimation of the convex combination parameters. Lastly, the estimation procedure requires development of an approximation to the Jacobian of the transformation, since it cannot be precalculated as in traditional SAR models (or as in Debarsy and LeSage, 2018).

In Section 2, we set forth matrix expressions for the model and discuss MCMC estimation. We also provide specifics regarding a computationally efficient approach that relies on trace approximations and quadratic forms involving outer vectors of parameters and inner matrices of sample data. In addition, we present the joint posterior distribution of the spatial dependence parameters in the model; an efficient approach to calculate partial derivative estimates needed to interpret the model; a Metropolis-Hastings tuned Monte Carlo integration approach to calculating the log-marginal likelihood; and our approach to

Bayesian model averaging which we compare to alternative estimation procedures.

An online appendix provides computational timing results as well as results from Monte Carlo experiments that explores bias, dispersion and coverage to validate our estimation procedure. An applied illustration of the method in a hedonic house price model is the subject of section 3. Finally, section 4 concludes.

## 2 Computationally efficient expressions for the model

The spatial autoregressive (SAR) model we wish to estimate is shown in (1), where each  $W_\ell$  represents an  $n \times n$  connectivity matrix whose main diagonal contains zero elements and row-sums of the off-diagonal elements equal one, with  $n$  being the number of observations. Non-zero (off-diagonal) matrix elements  $i, j$  of each  $W_\ell$  reflect that observation  $i$  is linked to observation  $j$ , with different connectivity matrices describing different possible types of interaction (e.g., geographic, economic, and so on).

$$\begin{aligned} y &= \rho W_c(\Gamma)y + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \\ W_c(\Gamma) &= \sum_{\ell=1}^L \gamma_\ell W_\ell, \quad 0 \leq \gamma_\ell \leq 1, \quad \sum_{\ell=1}^L \gamma_\ell = 1, \\ \Gamma &= (\gamma_1, \dots, \gamma_L)' \end{aligned} \tag{1}$$

The  $n \times k$  matrix  $X$  in (1) contains exogenous explanatory variables, with  $\beta$  being the associated  $k \times 1$  vector of parameters. The  $n \times 1$  vector  $\varepsilon$  represents a constant variance normally distributed disturbance term.

The SAR model in (1) can be re-expressed as in (2), a computationally convenient

expression that isolates the parameters  $\rho, \gamma_\ell, \ell = 1, \dots, L$  in the  $(L + 1) \times 1$  vector  $\omega$ .

$$\begin{aligned}\tilde{y}\omega &= X\beta + \varepsilon \\ \tilde{y} &= (y, W_1y, W_2y, \dots, W_Ly) \\ \omega &= (1, -\rho\Gamma')'.\end{aligned}\tag{2}$$

The value of isolating the parameter vector  $\omega$  is that we can pre-calculate the  $n \times (L + 1)$  matrix  $\tilde{y}$  prior to the beginning of the MCMC sampling loop, which only contains sample data.

The likelihood of the model in (2) is shown in (3), where  $\mathcal{W} = W_1, \dots, W_L$ .

$$\begin{aligned}f(y|X, \mathcal{W}; \rho, \Gamma, \sigma^2, \beta) &= |R(\omega)| (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{e'e}{2\sigma^2}\right) \\ e &= \tilde{y}\omega - X\beta \\ R(\omega) &= I_n - \rho W_c(\Gamma)\end{aligned}\tag{3}$$

where  $|R(\omega)|$  is the Jacobian of the transformation, which in this case depends on the parameters  $(\rho, \Gamma)$ . We restrict  $\rho \in (-1, 1)$ , so that  $R(\omega)^{-1} = \sum_{j=0}^{\infty} \rho^j W_c^j(\Gamma)$  exhibits an underlying stationary process. Some authors use  $\rho \in (-\lambda_n^{-1}, 1)$ , where  $\lambda_n$  is the smallest eigenvalue of  $W$  and the upper bound of 1 arises as the maximum eigenvalue of row-normalized matrices  $W$ . By considering  $(-1, 1)$  as parameter space for  $\rho$ , we avoid computing the minimum eigenvalue of  $W_c(\Gamma)$ . The parameter space for the set of parameters  $(\rho, \Gamma, \sigma^2, \beta)$  is:  $\Omega := \Omega_\rho \times \Omega_\Gamma \times \Omega_\sigma \times \Omega_\beta = (-1, 1) \times [0, 1]^L \times (0, \infty) \times \mathcal{R}^k$ .

## 2.1 Flat priors for the parameters

Mathematically, the flat or uniform priors for  $\rho, \Gamma, \beta$  can be represented as  $p(\rho) \sim U(-1, 1)$ ,  $p(\Gamma) \sim U(0, 1)$  and  $p(\beta) \propto 1$ . The noise variance  $\sigma^2$ , is restricted to positive values. It is customary to apply a flat prior to the log-transformed value,  $\log(\sigma^2)$ . By retransforming

the flat prior for  $\log(\sigma^2)$  in terms of the original noise variance, we get  $p(\sigma^2) \propto 1/\sigma^2$  (see Dittrich et al., 2017, p.216). Given this prior information, and prior independence, we can write:  $p(\rho) \times p(\Gamma) \times p(\sigma^2) \times p(\beta) \propto 1/\sigma^2$ , where the uniform priors for the parameters  $\rho$  and  $\Gamma$  reflect proper (bounded) probability distributions, while the priors on  $\sigma^2$  and  $\beta$  are unbounded and improper since the integral over their parameter space ( $\Omega_{\sigma^2}$  and  $\Omega_{\beta}$ ) is not finite. In the context of alternative models involving different weight matrices, if we rely on the same uniform priors for  $\rho$ ,  $\Gamma$  and integrate out the parameters  $\beta$  and  $\sigma^2$ , Hepple (1995*a,b*) has shown that the joint posterior distribution for the cross-sectional dependence parameters  $\rho$  and  $\Gamma$  are proper under relatively unrestrictive assumptions.

Combining the likelihood function in (3) with the flat priors (and ignoring the constant term  $2\pi^{n/2}$ ) leads to the joint posterior  $p(\rho, \Gamma, \beta, \sigma^2)$  in (4), from which  $\sigma$  can be integrated out, leading to (6).

$$p(\rho, \Gamma, \sigma^2, \beta | y, X, \mathcal{W}) \propto |R(\omega)| (\sigma^2)^{-\frac{(n+2)}{2}} \exp\left(-\frac{1}{2\sigma^2} e'e\right) \quad (4)$$

$$e = \tilde{y}\omega - X\beta$$

$$p(\rho, \Gamma, \beta | y, X, \mathcal{W}) \propto |R(\omega)| \int_0^\infty \sigma^{-(n+2)} \exp\left(-\frac{1}{2\sigma^2} e'e\right) d\sigma \quad (5)$$

$$\propto |R(\omega)| (e'e)^{-n/2} \quad (6)$$

To integrate out the  $k$  different  $\beta$  parameters, properties of the multivariate  $t$ -distribution in conjunction with ‘completing the square’ are used (see Zellner, 1971). This leads to a joint distribution for the dependence parameters  $\omega$  shown in (7), with the term  $|X'X|^{-1/2}$  and the exponent  $-(n-k)/2$  arising from this integration (see Hepple, 1995*a,b*). This expression must be numerically integrated with respect to  $\omega$  to arrive at the log-marginal likelihood for these models. This is accomplished using Monte Carlo integration discussed

later.

$$\begin{aligned}
 p(\omega|\tilde{y}, X, \mathcal{W}) &\propto |R(\omega)| |X'X|^{-1/2} (\omega'F\omega)^{-(n-k)/2} & (7) \\
 F &= U'U \\
 U &= M\tilde{y}, \quad M = I_n - X(X'X)^{-1}X'
 \end{aligned}$$

A question that has been explored in the literature is whether this conditional posterior distribution is proper and can be integrated over the parameter space for the case of the flat priors described here. Dittrich et al. (2017) tackle the traditional SAR model (relying on a single weight matrix  $W$ ) and conclude that propriety of this distribution requires conditions that extend to our model since  $W_c = \gamma_1 W_1 + \dots + \gamma_L W_L$  is equivalent to the single weight matrix  $W$ , given our restrictions on the parameters  $\gamma_\ell, \ell = 1, \dots, L$ .

## 2.2 The Markov Chain Monte Carlo (MCMC) estimation scheme

Successful estimation of parameters for the model in (1) requires a sufficiently large sample  $n$  of observations. To see this, note that the matrices  $W_\ell$  reflect important sample data in this type of model, as we wish to make distinctions between alternative specifications of  $W_\ell$ . Highly correlated connectivity matrices will lead to problems identifying the parameters  $\Gamma$ . Distinguishing between alternative interaction structures also requires that the dependence reflected by the parameter  $\rho$  is different from zero, which should be clear when considering that for  $\rho = 0$ , the parameters  $\Gamma$  are not identifiable. Since our focus is on large samples  $n$ , reliance on the flat and uniform priors will not likely impact posterior estimates.

The conditional distribution for the parameters  $\beta$  is multivariate normal with mean and



variance-covariance shown in (8).

$$\begin{aligned}
p(\beta|\sigma^2, \omega, \tilde{y}, X) &= N(\tilde{\beta}, \tilde{\Sigma}_\beta) \\
\tilde{\beta} &= (X'X)^{-1}(X'\tilde{y}\omega) \\
\tilde{\Sigma}_\beta &= \sigma^2(X'X)^{-1}
\end{aligned} \tag{8}$$

The conditional posterior for  $\sigma^2$  (given  $\tilde{\beta}, \omega$ ) takes the form in (9), given the prior  $p(\sigma^2) \propto 1/\sigma^2$ .

$$\begin{aligned}
p(\sigma^2|\beta, \omega, \tilde{y}, X) &\propto (\sigma^2)^{-\left(\frac{n}{2}\right)} \exp\left(-\frac{e'e}{2\sigma^2}\right) \\
e &= \tilde{y}\omega - X\beta \\
&\sim IG(\tilde{a}, \tilde{b}) \\
\tilde{a} &= n/2 \\
\tilde{b} &= (e'e)/2
\end{aligned} \tag{9}$$

The (log) joint posterior for the parameters in  $\omega$  after integrating out the parameters  $\beta$  and  $\sigma^2$  takes the form in (10) and does not reflect a known distribution (as in the case of the conditional distributions for  $\beta$  and  $\sigma^2$ ).

$$\ln p(\omega|\tilde{y}, X, \mathcal{W}) \propto \ln |R(\omega)| - \frac{n-k}{2} \ln(\omega'F\omega) \tag{10}$$

We note that  $\ln(\omega'F\omega)$  is a quadratic form involving the matrix  $F$ , which consists of only sample data and can be calculated prior to MCMC sampling. As such  $\ln(\omega'F\omega)$  can be quickly evaluated for any vector of dependence parameters  $\omega$ .

Finally, we sample the parameter  $\rho$  from the joint posterior distribution in (10) conditioning on  $\Gamma$  and similarly for  $\Gamma$  conditioning on  $\rho$ . Details about the sampling procedure are postponed to section 2.4.

### 2.3 Log-determinants based on trace approximations

To approximate the logarithm of the Jacobian, we rely on a (truncated) Taylor series expansion, shown in (11), which has been initially set forth by Pace et al. (2002) and Golub and Van Loan (1996). We use a fourth order Taylor expansion which Monte Carlo simulations reported in the online Appendix show works well to produce accurate parameter estimates as well as good coverage intervals.

$$\ln|I_n - \rho W_c(\Gamma)| = - \sum_{i=1}^{\infty} \rho^i \text{tr}(W_c^i(\Gamma))/i \quad (11)$$

$$\simeq - \sum_{j=2}^q \rho^j \text{tr}(W_c^j(\Gamma))/j \quad (12)$$

Note that the 1st-order  $\text{tr}(W_c(\Gamma))$  is zero, given the definitions of  $W_\ell$  and the restrictions placed on  $\gamma_\ell$ . The second-order trace can be expressed as a quadratic form in parameters  $\Gamma$  and all pairwise multiplications of the individual matrices in  $W_\ell$  as shown in (14) and (15), which are equivalent to expression (13).

$$\text{tr}(W_c^2(\Gamma)) = \sum_{i=1}^L \sum_{j=1}^L \gamma_i \gamma_j \text{tr}(W_i W_j) \quad (13)$$

$$= \Gamma' Q \Gamma. \quad (14)$$

$$= (\Gamma \otimes \Gamma)' \text{vec}(Q), \quad (15)$$

$$Q = \begin{pmatrix} \text{tr}(W_1^2) & \text{tr}(W_1 W_2) & \dots & \text{tr}(W_1 W_L) \\ \text{tr}(W_2 W_1) & \text{tr}(W_2^2) & \dots & \text{tr}(W_2 W_L) \\ \vdots & & \ddots & \\ \text{tr}(W_L W_1) & \text{tr}(W_L W_2) & \dots & \text{tr}(W_L^2) \end{pmatrix}. \quad (16)$$

LeSage and Pace (2009) point out that accelerated computation of traces can be accomplished using sums of matrix Hadamard products, which for *asymmetric* nonnegative connectivity matrices take the following form:  $\text{tr}(W_i W_j) = \sum_i^L \sum_j^L W_i \odot W_j'$ .

Expression (14) shows that the parameters can be separated from the trace of the product of matrices, which can be pre-calculated prior to MCMC sampling. An even more efficient computational expression to compute  $\text{tr}(W_c^2)$  is shown in (15), where  $\otimes$  stands for the Kronecker product and  $\text{vec}(Q)$  is the operator which stacks the columns of the matrix  $Q$ .

Expressions for 3rd- and 4th-order traces are presented in (17) and (18) respectively.

$$\text{tr}(W_c^3(\Gamma)) = \sum_{i=1}^L \sum_{j=1}^L \sum_{k=1}^L \gamma_i \gamma_j \gamma_k \text{tr}(W_i W_j W_k), \quad (17)$$

$$\text{tr}(W_c^4(\Gamma)) = \sum_{i=1}^L \sum_{j=1}^L \sum_{k=1}^L \sum_{l=1}^L \gamma_i \gamma_j \gamma_k \gamma_l \text{tr}(W_i W_j W_k W_l). \quad (18)$$

## 2.4 Sampling procedures for $\rho$ and $\Gamma$

As noted in the introduction, a second computational challenge for MCMC estimation of the model is sampling parameters in the vector  $\Gamma$ , which must sum to one and cannot be negative. We set forth a block sampling procedure for  $\Gamma$  that proposes a vector of candidate values for  $\gamma_\ell$ ,  $\ell = 1, 2, \dots, L-1$ , with  $\gamma_L = 1 - \sum_{\ell=1}^{L-1} \gamma_\ell$ . The conditional distributions for the current and proposed vectors that we label  $\Gamma^c$  and  $\Gamma^p$  are evaluated with a M-H step used to either accept or reject the newly proposed vector  $\Gamma^p$ . Block sampling the parameter vector  $\Gamma$  has the virtue that accepted vectors will obey the summing up restriction and reduce autocorrelation in the MCMC draws for these parameters. However, block sampling is known to produce lower acceptance rates which may require more MCMC draws in order to collect a sufficiently large sample of draws for posterior inference regarding  $\Gamma$ .

The block sampling procedure involves (for each  $\gamma_\ell$ ,  $\ell = 1, \dots, L-1$ ) a three-headed coin flip. By this we mean a uniform random number on the closed interval  $\text{coin flip} = U(0, 1)$ , with head #1 a value of  $\text{coin flip} \leq 1/3$ , head #2 a value in  $(1/3, 2/3]$  and head #3 a value  $> 2/3$ . Given a head #1 result, we set a proposal for  $\gamma_\ell^p$  using a uniform random draw on the interval  $[0, \gamma_\ell^c)$ , the current value. A head #2 results in setting the proposal

value equal to the current value ( $\gamma_\ell^p = \gamma_\ell^c$ ), while a head #3 selects a proposal value based on a uniform random draw on the interval  $(\gamma_\ell^c, 1]$ .

The (non-logged) conditional distributions in expression (19) are used to calculate a M-H acceptance probability.

$$\psi_{MH}(\Gamma^c, \Gamma^p) = \min(1, \exp[(\ln p(\Gamma^p|\rho) - \ln p(\Gamma^c|\rho))]) \quad (19)$$

The expression to be evaluated at the current and proposed vectors of parameters  $\Gamma$  consists of two relevant terms, one involving the log-determinant and the other the quadratic form:  $\ln(\omega'F\omega)$ , both of these evaluated for the vector of parameters  $\Gamma$ . As already motivated, our fourth-order Taylor series approximation to the log-determinant  $\ln|I_n - \rho W_c(\Gamma)|$  can be easily and rapidly calculated for any vector  $\Gamma$  using the pre-calculated traces and the conditioning parameter  $\rho$ , and use of these traces allows us to avoid the need to calculate the  $n \times n$  matrix  $W_c(\Gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$ , which saves on computer memory. A second point is that the quadratic form expression  $\ln[\omega'F\omega]$  can be easily calculated using the pre-calculated expression  $\tilde{y} = (y W_1 y W_2 y \dots W_L y)$  and the vector  $\omega$ .

After the initial 1,000 draws we switch to a tuned random-walk proposal procedure. This is needed because proposals from the block sampling procedure based on the large intervals between  $[0, \gamma_j^c)$  and  $(\gamma_j^c, 1]$  will not produce candidates likely to be accepted when these parameters are estimated with a great deal of precision, as would be the case for problems involving large sample size. This can result in a failure to move the chain adequately over the parameter space. To address this issue, standard deviations,  $\sigma_{\gamma(j)}$  for each parameter  $j = 1, \dots, L-1$  are calculated based on the first 1,000 draws (and thereafter using a rolling window interval of 1,000 draws). These are used in a tuned random-walk procedure to produce candidate/proposal values. Specifically, we use a tuning scalar that is adjusted based on acceptance rates for the block of parameters  $\Gamma$ . This is used in conjunction with the standard deviations and the block sampling procedure described previously to produce

proposals. Monte Carlo experiments in the online Appendix demonstrate the efficacy of this approach to sampling the parameter vector  $\Gamma$ .

Finally, the parameter  $\rho$  is drawn from the joint distribution for  $\omega$  shown in (10) conditioning on  $\Gamma$ . We follow LeSage and Pace (2009) and use M-H sampling for this parameter, based on a normal distribution along with a tuned random-walk procedure to produce candidate values for  $\rho$ . We further rely on rejection sampling to guarantee that proposed values belong to the  $(-1, 1)$  parameter space.

## 2.5 Calculating effects estimates

In addition to producing estimates for the underlying model parameters  $\rho, \beta, \Gamma, \sigma^2$ , the non-linear nature of the model relationship requires calculating partial derivatives of  $y$  with respect to changes in the exogenous variables  $X$  based on the reduced form of the model. For the standard SAR model, LeSage and Pace (2009) derive these partial derivatives along with scalar summary measures reflecting own- and cross-partial derivatives that they label *average direct* and *indirect* effects. For the SAR model with a convex combination of connectivity matrices, the partial derivative of  $y$  with respect to the  $r^{th}$  explanatory variable  $X^r$  is shown in (20) while the associated scalar summary measures are presented

in (21) to (25).

$$\partial y / \partial X^r = S_r(W_c(\Gamma)) \quad (20)$$

$$S_r(W_c(\Gamma)) = (I_n - \rho W_c(\Gamma))^{-1} \beta_r$$

$$\bar{M}(r)_{direct} = n^{-1} \text{tr}(S_r(W_c(\Gamma))) \quad (21)$$

$$= n^{-1} \beta_r \sum_{i=0}^{\infty} \rho^i \text{tr}(W_c^i(\Gamma)) \quad (22)$$

$$\approx n^{-1} \beta_r \sum_{i=0}^q \rho^i \text{tr}(W_c^i(\Gamma)) \quad (23)$$

$$\bar{M}(r)_{total} = \frac{\beta_r}{(1 - \rho)} \quad (24)$$

$$\bar{M}(r)_{indirect} = \bar{M}(r)_{total} - \bar{M}(r)_{direct} \quad (25)$$

While expressions in (21), (24) and (25) produce point estimates for the scalar summary measures of effects, we also require measures of dispersion for statistical inference regarding the significance of these effects. Use of an empirical distribution constructed by simulating the non-linear expressions in (20) using (say 1,000) draws from the posterior distribution of the underlying parameters  $\rho, \beta_r, \gamma_\ell, \ell = 1, \dots, L$  is suggested by LeSage and Pace (2009).

Expression (22) shows that the only required quantities for constructing scalar effect measures are the traces of powers of  $W_c(\Gamma)$ . In practice, this infinite sum is truncated, as indicated by (23). LeSage and Pace (2009), who initially set forth this computationally efficient solution rely on  $q = 100$ . See LeSage and Pace (2009, chap.4) for details regarding computation of measures of dispersion for the scalar summary effects estimates in the traditional SAR model.

The computation of the traces in (23) is complicated by the fact that the matrix  $W_c(\Gamma)$  depends on estimated parameters  $\gamma_\ell, \ell = 1, \dots, L$  ruling out use of pre-calculated traces. To account for the stochastic variation in the effects estimated due to uncertainty in the estimation of the parameter  $\Gamma$ , we proceed in the following steps.

We calculate  $q = 100$  traces using  $W_c(\bar{\Gamma}) = \sum_{\ell=1}^L \bar{\gamma}_\ell W_\ell$  where  $\bar{\Gamma}$  is the posterior mean

value of the vector  $\Gamma$  (more precisely, the traces are not computed but estimated using the procedure set forth in LeSage and Pace, 2009). The second- through fourth-order traces are then replaced with those shown in (13), (17) and (18) during simulation, where MCMC draws for the parameters  $\gamma_\ell$  are used instead of relying on the posterior mean value of these parameters.

Given that the sum in (23) involves increasingly small magnitudes associated with higher-order powers of the parameters  $\rho$  and  $\Gamma$ , low-order traces are most important for accurate estimates of the effects. With the exception of modeling situations involving very large negative or positive values of spatial dependence  $\rho$  (say greater than 0.9 in absolute value) which do not arise often in applied practice, this approach should produce accurate estimates for the effects, as shown in the Monte Carlo experiments reported in the online Appendix.

## 2.6 Accounting for model uncertainty

The convex combination of connectivity matrices model specification raises the question of which matrices should be used and which should be ignored. For example, in the case of  $L$  candidate connectivity matrices, there are  $M = 2^L - L - 1$  possible ways to employ *two or more* of the  $L$  matrices in alternative model specifications. When  $L = 5$ , we have  $M = 26$  possible models involving two or more matrices, and for  $L = 10$ ,  $M = 1,013$ . To account for this model uncertainty, we rely on Metropolis-Hastings guided Monte Carlo integration during MCMC estimation of the models to produce log-marginal likelihoods and associated posterior model probabilities for the set of  $M$  possible models, which allows for Bayesian model averaged estimates. Alternative Bayesian procedures developed in the literature to account for model uncertainty are also discussed.

### 2.6.1 Computation of log-marginal likelihoods

Calculation of the log marginal likelihood for model (1) involves integrating the joint posterior distribution over all model parameters. We can analytically integrate out the parameters  $\beta$  and  $\sigma^2$ , leading to a (log kernel) joint posterior for the remaining model parameters in  $\omega$  that takes the form in (10), (see Hepple, 1995*a,b*).

Integrating out the remaining parameters using numerical methods could be computationally intensive for cases involving a large number of parameters in the vector  $\Gamma$ . We rely on Metropolis-Hastings tuned Monte Carlo integration that avoids the efficiency problem typically encountered with Monte Carlo integration when samples would reflect areas of low support in the joint posterior distribution. Use of MCMC draws for the parameters  $\omega$  based on the Metropolis-Hastings sampler makes the integration very efficient because these are located in areas of high density in the posterior distribution. The Monte Carlo integration estimate of the the (kernel) joint posterior is the mean of expression (10) evaluated for values in the vector  $\omega$  on every trip through the MCMC sampling loop.

We can then add any constants of integration to produce an estimate of the log marginal likelihood. These constants take the form:  $\kappa = -\log(1/\max \rho - 1/\min \rho) + \log(\text{Gamma}(dof)) - dof \times \log(2\pi)$ , where  $\log(\text{Gamma}())$  is the log gamma function, and  $dof$  are the degrees of freedom equal to  $(n - L)/2$ , where we loose degrees of freedom based on the number of weight matrices. The (normalized) non-log joint posterior can be calculated in the usual way using  $\exp[\log p(\omega) - \max(\log p(\omega))]$ .

### 2.6.2 Bayesian model averaging

Given the speed of estimation for single models and the availability of multi-core computer architecture, it is possible to estimate models based on all possible combinations of two or more connectivity matrices, even in cases of 10 matrices.

During estimation, log-marginal likelihood estimates would be produced that allow



calculation of posterior model probabilities for the set of  $M$  models.

Given the non-linear relationship between the underlying parameters  $\beta, \Gamma, \rho$  and the scalar summary measures of direct and indirect effects which are the focus of inference in these models, model averaged estimates should be constructed by applying model probabilities to the scalar summary estimates of the direct and indirect effects from each model.

As an illustration, we present estimation results for all models ( $M = 26$ ) involving two or more connectivity matrices using a set of five candidate  $W$ -matrices and a sample of  $N = 2,000$  observations in Table 1. The DGP in (26) was used where the three first candidate matrices have non-zero gamma values,  $\gamma_1 = 0.4, \gamma_2 = 0.3, \gamma_3 = 0.3$ . We set  $\gamma_4 = \gamma_5 = 0$ , as a test of overfitting for this type of convex combination.

$$\begin{aligned} y &= \rho(\gamma_1 W_1 + \gamma_2 W_2 + \gamma_3 W_3)y + X\beta + \varepsilon \\ &= (I_n - \rho W_c(\Gamma))^{-1}(X\beta + \varepsilon) \end{aligned} \tag{26}$$

Where  $X$  includes a constant term and 2 standard normal variables,  $\beta = (-1, -0.5, 1.5)'$  and  $\rho = 0.6$ . The error term  $\varepsilon$  is assumed normally distributed, centered around zero with a variance  $\sigma^2 I_n$  where  $\sigma^2 = 3.6061$  so the signal to noise ratio ( $SNR$ ) of the model is defined following Debarsy and LeSage (2018) to equal to 0.7. They define SNR as in (27),

$$SNR = \frac{A'A}{A'A + \sigma^2 tr(B'B)} \tag{27}$$

where  $A = (I_n - \rho W_c(\Gamma))^{-1}X\beta$  and  $B = (I_n - \rho W_c(\Gamma))^{-1}$ .

The five candidate  $W$ -matrices are based on five nearest neighbors matrices constructed from independent sets of random normal latitude-longitude vectors to ensure they convey different connectivity information. A set of 60,000 draws were used with the first 10,000 discarded for burn-in, and thinning of the 50,000 retained draws was used based on every fifth draw producing a sample of 10,000 draws used to construct estimates.

Estimates for  $\rho$  and the parameters  $\gamma_\ell, \ell = 1, \dots, 5$  for the 26 possible models involving

combinations of 2 or more of the five  $W$ -matrices are shown in the upper part of Table 1, along with the log-marginal likelihood estimate and posterior model probabilities. From the table, we see that the true model (model #11) that generated the sample data vector  $y$  had a posterior model probability of 92.19%, with model #21 assigned a probability of 3.68% and model #22 a probability of 3.96%. Note that model #21 contains the three true matrices  $W_1, W_2, W_3$  plus  $W_4$  with a  $\hat{\gamma}_4 = 0.0366$ , and model #22 contains the three true matrices  $W_1, W_2, W_3$  plus  $W_5$  with a  $\hat{\gamma}_5 = 0.0415$ , both of which are plausible models to receive some support in terms of posterior model probabilities.

Table 1 also shows posterior means for a set of model averaged estimates for the parameters  $\rho, \gamma_\ell, \ell = 1, \dots, 5$ . These were constructed using the model probabilities to weight the 10,000 retained MCMC draws from each of the 26 models, with posterior means calculated based on the set of 10,000 probability-weighted draws. We see small values for the BMA parameters  $\hat{\gamma}_4 = 0.0014, \hat{\gamma}_5 = 0.0017$ . We can calculate credible intervals for these two parameters using the distribution of the 10,000 probability-weighted draws, which would allow us to determine if estimates for model #11 suffers from problems with inference at the boundary of the parameter space for  $\Gamma$ . That is, we can use credible intervals for the BMA posterior estimates of  $\gamma_4, \gamma_5$  to see if an inference of zero for these parameters is reasonable. The 0.01 interval for  $\hat{\gamma}_4$  is 0.0046, and that for  $\hat{\gamma}_5$  is 0.005, allowing us to conclude that these parameters are sufficiently close to zero to make an inference of zero reasonable.

### 2.6.3 Alternative estimation methods

There is a literature on Bayesian model averaging for spatial autoregressive models, where the focus has been on extending the approaches of Fernández et al. (2001) and George and McCulloch (1993, 1997) from non-spatial regression modeling that focuses on selection of explanatory variables in the matrix  $X$ . LeSage and Parent (2007) show how to extend the approach of Fernández et al. (2001) that relies on the log-marginal likelihood for models consisting of differing explanatory variables using univariate numerical integration of the

spatial dependence parameter that arises in the spatial autoregressive model (with a single weight matrix) to arrive at the log-marginal likelihood. They rely on the same reversible jump MCMC estimation approach as Fernández et al. (2001) to explore alternative explanatory variables as well as the Zellner (1986)  $g$ -prior distribution for the coefficients association with these variables, while fixing the single spatial weight matrix. LeSage and Fischer (2008) extend this approach slightly to sample alternative (single) weight matrices based on differing numbers of nearest neighbors.

Although LeSage and Fischer (2008) allow for uncertainty in the connectivity matrix, the SAR model they consider allows for only one connectivity matrix. In other words, they implicitly assume (as does most of the spatial econometrics literature) that the cross-sectional dependence scheme is composed of only one layer. Our approach allows for a more general interaction scheme based on a convex combination of several simultaneously interacting layers. We note that estimating a SAR model with a single connectivity matrix when the DGP involves several of these matrices should lead to unreliable estimates. To illustrate this point, we perform a small Monte Carlo analysis where we compare the LeSage and Fischer (2008) approach with the one developed here.

In this setup, we consider uncertainty solely in the specification of the connectivity matrix and for the sake of simplicity, we only consider two possible choices:  $W_1$  and  $W_2$ .

We compare the LeSage and Fischer (2008) approach with the one developed here under two alternative DGPs. The first DGP is based on the cross-sectional dependence scheme involving a convex combination of the two matrices  $W_1$  and  $W_2$  as shown in (28), while the second uses only the matrix  $W_1$  in the DGP as shown in (29), where  $\iota_n$  is an  $n \times 1$  vector of ones.

$$y = \rho(\gamma_1 W_1 + (1 - \gamma_1) W_2)y + \beta_0 \iota_n + X_1 \beta_1 + \varepsilon \quad (28)$$

$$y = \rho W_1 y + \beta_0 \iota_n + X_1 \beta_1 + \varepsilon \quad (29)$$

Matrices  $W_1$  and  $W_2$  are based on five nearest neighbors, where each connectivity matrix

was generated using different sets of standard normal latitude-longitude coordinates. We consider a standard normal variable for the vector  $X_1$  (kept constant across the 1000 replications) and also for the error term. Finally, the sample size is set to 2000 observations.

To compare the two approaches, we produce estimates based on either the DGP from (28) or (29) using matrices  $W_1$  and  $W_2$  separately, and then calculate Bayesian model averaging estimates based on the two sets of results (approach of LeSage and Fischer, 2008). We contrast these BMA results with those obtained from the convex combination methodology set forth here.

The first column of Tables 2 to 4 shows the true value of the parameters used in the DGP. Columns 2 to 4 summarize the results obtained with LeSage and Fischer (2008). The last column reports the results based on the convex combination approach. Table 2 reports the results based on the DGP with a convex combination of connectivity matrices, setting  $\gamma_1 = 0.5$ . As expected, the convex combination approach performs much better. Columns 2 to 4 show substantial bias in the estimated coefficients and associated effects estimates arising from model misspecification of the connectivity matrix. Large bias in the estimate of  $\rho$  leads to direct, indirect and total effects far from their true value. (We note that these effects estimates are the focus of interest and the basis for inference in SAR models.) Since  $\gamma_1 = 0.5$ , the posterior probability is roughly the same for each model based on a single connectivity matrix, as we would expect.

Table 3 summarizes the results when  $\gamma_1 = 0.2$ , with findings similar to those from Table 2. We note that the SAR model with only  $W_2$  is always chosen over that based on  $W_1$  only. Results based on the convex combination approach are always much better.

Finally, Table 4 reports results for a case where the DGP is a SAR based on  $W_1$  only (equation 29) using the same format as in preceding tables. In this case, the convex combination performs well, but remains slightly inferior to estimates obtained from the true DGP model. This is expected since the estimated parameter governing the convex combination ( $\gamma_1$ ) will not exactly equal that implied by the true DGP ( $\gamma_1 = 1$ ).

A second alternative strategy has been proposed by Piribauer and Crespo Cuaresma (2016) who focus on a single weight matrix for the spatial autoregressive model constructed based on some unknown number of nearest neighboring regions and adopt the stochastic variable selection approach from George and McCulloch (1997). This approach introduces a parameter that controls inclusion of explanatory variables in the model, and uses the posterior inclusion probability for this parameter as a way to draw inferences regarding which variables are important. Piribauer and Crespo Cuaresma (2016) extend the parameter that controls inclusion of explanatory variables in the model to also govern inclusion of a weight matrix based on some variable number of nearest neighbors. This allows a posterior inference regarding the number of neighbors that are most appropriate for a single weight matrix. They argue that applying the stochastic variable selection approach used for selection of explanatory variables to also determine the number of nearest neighbors weight matrix in spatial autoregressive models holds computational advantages over the approach taken by LeSage and Fischer (2008). In the applied illustration presented in the next section, we show that for the case of a single connectivity matrix, the two estimation strategies (LeSage and Fischer, 2008; Piribauer and Crespo Cuaresma, 2016) find the same optimal matrix. However, the same drawback illustrated above for the LeSage and Fischer (2008) approach applies to the Piribauer and Crespo Cuaresma (2016) approach in cases where the true DGP involves convex combinations of weight matrices.

### **3 An applied illustration**

To illustrate the method, we estimate a hedonic house price regression using a sample of 72,045 homes sold in the state of Ohio during the year 2000. The data is described in Brasington and Haurin (2006); Brasington (2007) and Brasington and Hite (2008). The dataset is publicly available at: <http://homepages.uc.edu/~brasindd/housing.html>. We rely on a subset from a larger sample of 112,830 homes, because we require that each

home has twenty nearest neighboring homes within two miles that had the the same number of bedrooms, same number of full plus half baths and same categorical age variable. Full plus half baths were assigned a value equal to the # of full + 0.5 x # of half-baths. So, a home with one full plus two half baths has a value of 2 (equal to two full baths), and so on. House age was specified as six categorical variables, 0 to  $\leq 5$  years (new homes up to those less than or equal to 5 years old), 6 to  $\leq 10$  years, 11 to  $\leq 20$  years, 21 to  $\leq 50$  years, 51 to  $\leq 100$  years, and more than 100 years.

The dependent variable is the (*logged*) *selling price*, with two explanatory variables:  $\log(\text{total living area})$  in square feet, denoted as *lTLA* and  $\log(\text{lot size})$  in square feet, which we label *lLSIZE*. Typically, house characteristics such as bedrooms, baths and house age are used as explanatory variables in hedonic house price regressions. We take a different approach and use these variables to specify a set of alternative connectivity matrices. The motivation for this approach is that the conventional spatial autoregressive hedonic house price regression uses a spatial lag of prices from nearby homes (whatever their characteristics) as a way to approximate prices of comparable homes, and treats house characteristics such as bedrooms, baths and house age as explanatory variables that explain variation in selling prices.

Our use of these characteristics to form a convex combination of connectivity matrices, treats the characteristics as reflecting a more composite notion that we label *house design*. House design is the basis on which a buyer searching for a home identifies comparable homes. Since the conventional spatial lag of nearby home selling prices is an approximation used to identify comparable homes on which selling price of each home is dependent, our approach could be viewed as an attempt to improve on identifying comparable homes on which selling price of each home is dependent. As such, we expect to see a higher level of cross-sectional dependence between selling prices. The higher level of dependence also implies that we would see larger indirect/spillover effects for the model that uses connectivity matrices reflecting design aspects of homes. In addition, we might see a denser weight

matrix arising from the convex combination approach, which would also lead to larger spatial spillovers, because these are calculated as an average of the cumulative off-diagonal elements of the  $n \times n$  matrix of partial derivatives, which are used to produce the scalar summary measures of indirect effects. In the following, we will compare our proposed approach to constructing the neighborhood of each home based on *house design* with the one based only on spatial neighbors without regard for house design. We also compare the convex combination approach based only on *ITLA* and *LSIZE* (in square feet) as explanatory variables with a more conventional specification that models cross-sectional dependence using only spatial neighbors and treats the number of bedrooms, baths and house age as explanatory variables in the hedonic house price relationship.

We note that one can view characteristics such as bedrooms, baths and house age as either explanatory variables (e.g., conventional hedonic house price relationship), or as facets of house design that define comparable homes (e.g., our convex combination model illustrated here), but not both. If the weight matrix used in a SAR model is functionally related to explanatory variables in the model, important interpretation difficulties arise. Consider that in this case, the partial derivative effect of a change in  $X^r$  on the outcome variable vector  $y$  would require that we take into account any (non-linear) change in the connectivity structure induced by the change in  $X^r$ . Lee and Yu (2012) develop theoretical expressions for this type of situation which require knowledge of the functional relationship between  $X^r$  and the connectivity matrix, something unlikely to be known in practice.

Three connectivity matrices were constructed to reflect the nearest neighboring homes that potential buyers would view as comparable along three dimensions. Specifically, the nearest homes (within two miles) with the same number of bedrooms ( $W_{beds}$ ), the same number of full plus half-baths ( $W_{baths}$ ), and the same age category ( $W_{age}$ ). In their search for a home, buyers would view nearby homes that did not have a similar number of beds, baths or house age as not truly comparable.

Of course, there is the question of how many nearest neighboring homes should be

used to construct each connectivity matrix (and hence the convex combination) which can be answered using estimated log-marginal likelihoods and associated model probabilities. Throughout the application, a set of 60,000 MCMC draws were made with the first 30,000 discarded for burn-in and thinning of the 30,000 retained draws was used to produce a sample of 1,000 draws used for inference. The model used to select the relevant number of neighbors for each definition of connectivity is shown in (30), where 4 types of connectedness were considered: three mentioned above as well as the conventional spatial connectivity matrix based on geographically nearest neighboring homes ( $W_{space}$ ).

$$y = \rho W y + \beta_0 t_n + \beta_1 lT L A + \beta_2 l L S I Z E + \varepsilon \quad (30)$$

Table 5 shows the results of these calculations while Table 6 reports the outcomes associated to the approach developed by Piribauer and Crespo Cuaresma (2016). We observe that both approaches lead to the same results, where the optimal number of neighbors for  $W_{space}$  is 14, 12 for  $W_{beds}$ , 13 for  $W_{baths}$  and 14 for  $W_{age}$ .

LeSage and Pace (2014) propose a measure of similarity for alternative connectivity matrices, that involves multiplication of each  $n \times n$  matrix with the same  $n \times 1$  random normal vector ( $\nu$ ) to produce vectors  $W_{space}\nu, W_{beds}\nu, W_{baths}\nu, W_{age}\nu$ . The correlation between these vectors can then be used to judge similarity. A correlation matrix based on the optimal number of neighbors is shown in Table 7, where we see correlations around 0.5 for lag vectors  $W_{beds}\nu$  and  $W_{baths}\nu$  and a slightly higher correlation of 0.54 between lag vectors  $W_{baths}\nu$  and  $W_{age}\nu$ . These correlations indicate that neighbors defined on the basis of homes with a common number of bedrooms, baths or of similar age are not the same. It is also of interest to consider the correlation between the three components that we use to define our notion of *house design* versus the conventional connectivity matrix based solely on nearby spatial neighbors. We see that the highest correlation between the spatial lag vector  $W_{space}\nu$  is with  $W_{age}\nu$ , indicating spatial clustering of homes of the same



vintage/age, which is not surprising given that homes in a neighborhood are usually built around the same time. These results point to the inherent truth that homes located nearby tend to be similar in design providing a good approximation to comparable homes, which accounts for the success of spatial autoregressive hedonic house price regressions that rely on prices of homes located nearby.

In this application, the convex combination of connectivity matrices is based on three different measures of similarity. Table 8 reports the log-marginal likelihood and associated posterior probabilities of all models that include two or more matrices in the convex combination. The results indicate a strong preference for a model based on the number of bedrooms, bathrooms and house age (model 4), since the posterior probability associated with this model is 100%. Table 9 reports estimation results for each connectivity matrix considered separately along with estimates based on the convex combination of house design.

Similar results in terms of estimated coefficients and log-marginal values are observed for the SAR model based on the four separate connectivity matrices, but we see some divergence in the calculated effects estimates. These differences are notable for the total and indirect elasticities of selling price with respect to total living area, which range between 1.261 and 1.499 for the former and between 0.919 and 1.077 for the latter.

Comparing results based on a conventional geographically based weight matrix with those from the convex combination model, we observe different findings in terms of the direct, indirect and total effects. Estimates for the parameters  $\gamma_1, \gamma_2, \gamma_3$  indicate that house age is the most important characteristic, with bathrooms next most important and bedrooms least important. Given the lower 0.01 and upper 0.99 credible intervals, these differences in the relative importance of house age, bathrooms and bedrooms are significant.

In addition, the lower 0.01 and upper 0.99 credible intervals point to a significant difference between the direct effects of TLA from the two models. Specifically, a 10 percent increase in living area (TLA) would result in a 4.2 percent higher price in the case of

the spatial neighbors specification, but only a 3.3 percent higher price for the convex combination model. Also, the spillover (indirect) effects associated with the 10 percent increase in TLA are larger for the convex combination model, implying a 12.46 percent increase in selling price versus an 10.77 percent higher price in the case of changes in TLA when the interaction scheme is solely based on geographic neighbors. These results accord with the motivation given earlier that better identification of comparable homes results in a higher level of dependence, with a median value  $\rho = 0.803$  for the convex combination model versus  $\rho = 0.739$  for the spatial neighbors model. Of course, this leads to larger indirect or spillover effects. As noted earlier, larger spillover estimates may also arise because the convex combination model weight matrix is denser than the spatial weight matrix. (A check of non-zero off-diagonal elements from the two weight matrices showed this was indeed the case. For example, the average number of non-zero off-diagonal elements of the convex combination matrix was 23, versus 14 for the spatial weight matrix based on 14 nearest neighbors.)

In the case of the elasticity response of selling price to changes in lotsize, we also observe a significant difference in the direct and indirect effects for the five cases under study. Here again, given the higher level of dependence for the convex combination model and the denser connectivity matrix  $W_c$ , we would expect to see a larger indirect effect of lotsize from more comparable homes on the selling prices. The significant difference in indirect lotsize effects estimates lead to a significant difference in the total effect of lotsize for the five specifications, with those from the convex combination model being larger.

It is also the case that the log-marginal likelihood shows a clear improvement for the convex combination model relative to that based on all proximity weights taken separately.

Finally, Table 10 compares the estimation results of the convex combination SAR model with house characteristics (number of bedrooms, baths and house age) used to construct the *house design* notion of comparable houses with those of a conventional SAR model where the neighborhood is based on a purely geographic 14 nearest neighbors definition

and where the characteristics are treated as explanatory variables.

We first note that the log-marginal likelihood value is much higher for the SAR model based on a convex combination of matrices. Adding the number of bedrooms, bathrooms and the house age to the set of explanatory variables slightly decreases estimated values of the coefficients and effects of *ITLA* and *LSIZE*, with respect to the first column of Table 9, dedicated to the estimation of the conventional SAR model where connectivity is based on the 14 geographic nearest neighbors, even though the log-marginal of the latter is somewhat smaller than for the former (-26,490.735 vs -26,052.668).

Finally, we observe that the approach relying on a convex combination of matrices implies much higher indirect and total effects for *ITLA* and *LSIZE* than the SAR model with the additional explanatory variables, even though direct effects are quite similar.

## 4 Conclusion

We consider estimating spatial regression models that utilize convex combinations of connectivity structures that expand on conventional approaches based only on a single connectivity scheme between observations. Models constructed using a convex combination of weight matrices to form a single linear combination of alternative weight structures hold intuitive appeal since cross-sectional dependence is most of the time multi-dimensional.

The spatial autoregressive model with convex combination of connectivity matrices raises the question of which matrices should be used and which should be ignored. We show how Metropolis-Hastings guided Monte Carlo integration can be used during MCMC estimation of the models to produce estimates of log-marginal likelihoods and associated posterior model probabilities for alternative models, which allows for Bayesian model averaged estimates.

An applied illustration of our model shows that spatial hedonic house price regressions that rely on the conventional spatial weight matrix that relates selling prices of homes to

those located nearby in space can be improved by taking into consideration design characteristics of homes located nearby. Homes located nearby of comparable age, bedrooms and baths can be viewed as comparable in terms of design, and better model the relevant neighborhood buyers (or real estate appraisers) consider to assess house prices.

## References

- Blankmeyer, E., LeSage, J. P., Stutzman, J. R., Knox, K. J. and Pace, R. K. (2011), ‘Peer-group dependence in salary benchmarking: a statistical model’, *Managerial and Decision Economics* **32**(2), 91–104.
- Brasington, D. (2007), ‘Private schools and the willingness to pay for public schooling’, *Education Finance and Policy* **2**, 152–174.
- Brasington, D. and Haurin, D. (2006), ‘Educational outcomes and house values: A test of the value added approach’, *Journal of Regional Science* **46**, 245–268.
- Brasington, D. and Hite, D. (2008), ‘A mixed index approach to identifying hedonic price models’, *Regional Science and Urban Economics* **38**, 271–284.
- Debarsy, N. and LeSage, J. (2018), ‘Flexible dependence modeling using convex combinations of different types of connectivity structures’, *Regional Science and Urban Economics* **69**, 48–68.
- Dittrich, D., Leenders, R. and Mulder, J. (2017), ‘Bayesian estimation of the network autocorrelation model’, *Social Networks* **48**, 213–236.
- Fernández, C., Ley, E. and Steel, M. (2001), ‘Benchmark priors for bayesian model averaging’, *Journal of Econometrics* **100**, 381–427.
- George, E. and McCulloch, R. (1993), ‘Variable selection via gibbs sampling’, *Journal of the American Statistical Association* **88**(423), 881–889.

- George, E. and McCulloch, R. (1997), ‘Approaches for bayesian variable selection’, *Statistica Sinica* **7**, 339–373.
- Golub, G. and Van Loan, C. (1996), *Matrix Computation, Third Edition*, The Johns Hopkins University Press, Baltimore.
- Hazır, C. S., LeSage, J. and Autant-Bernard, C. (2018), ‘The role of r&d collaboration networks on regional knowledge creation: Evidence from information and communication technologies’, *Papers in Regional Science* **97**(3), 549–567.
- Hepple, L. (1995a), ‘Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds’, *Environment and Planning A* **27**(3), 447–469.
- Hepple, L. (1995b), ‘Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms’, *Environment and Planning A* **27**(4), 615–644.
- Lee, L.-f. and Yu, J. (2012), ‘QML estimation of spatial dynamic panel data models with time varying spatial weights matrices’, *Spatial Economic Analysis* **7**, 31–74.
- LeSage, J. and Fischer, M. (2008), ‘Spatial growth regressions: model specification, estimation and interpretation’, *Spatial Economic Analysis* **3**(3), 275–304.
- LeSage, J. P. and Pace, K. (2014), ‘The biggest myth in spatial econometrics’, *Econometrics* **2**(4), 217–249.
- LeSage, J. and Pace, K. (2011), ‘Pitfalls in higher order model extensions of basic spatial regression methodology’, *The Review of Regional Studies* **41**(1), 13–26.
- LeSage, J. and Pace, K. R. (2009), *Introduction to Spatial Econometrics*, CRC Press Taylor and Francis Group, Boca Raton.
- LeSage, J. and Parent, O. (2007), ‘Bayesian model averaging for spatial econometric models’, *Geographical Analysis* **39**, 241–267.

- Pace, K., Barry, R., Gilley, O. and Sirmans, C. (2000), 'A method for spatial-temporal forecasting with an application to real estate prices', *International Journal of Forecasting* **16**, 229–246.
- Pace, K. R. and LeSage, J. (2002), 'Semiparametric maximum likelihood estimates of spatial dependence', *Geographical Analysis* **34**(1), 76–90.
- Pace, R., Sirmans, C. and Slawson, V. (2002), Are appraisers statisticians?, *in* K. Wang and M. Wolverton, eds, 'Real Estate Valuation Theory', Springer, pp. 31–45.
- Piribauer, P. and Crespo Cuaresma, J. (2016), 'Bayesian variable selection in spatial autoregressive models', *Spatial Economic Analysis* **11**, 457–479.
- Zellner, A. (1971), *Introduction to Bayesian inference in Econometrics*, John Wiley, New York.
- Zellner, A. (1986), On assessing prior distributions and bayesian regression analysis with  $g$ -prior distributions, *in* P. Goel and A. Zellner, eds, 'Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti', North-Holland/Elsevier, pp. 233–243.

Table 1: Estimates for 5 candidate  $W$ -matrices,  $M = 26$  models,  $N = 2000$ .

Models	log marginal	Prob( $m_i$ )	$\rho$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_5$
1	-4860.8213	0.0000	0.4201	0.5688	0.4312	-	-	-
2	-4866.3998	0.0000	0.3979	0.6023	-	0.3977	-	-
3	-4881.2037	0.0000	0.2568	0.9194	-	-	0.0806	-
4	-4881.1596	0.0000	0.2599	0.9096	-	-	-	0.0904
5	-4882.0701	0.0000	0.3353	-	0.5378	0.4622	-	-
6	-4896.0347	0.0000	0.2011	-	0.8924	-	0.1076	-
7	-4895.9268	0.0000	0.2042	-	0.8818	-	-	0.1182
8	-4901.7257	0.0000	0.1757	-	-	0.8746	0.1254	-
9	-4901.6148	0.0000	0.1799	-	-	0.8579	-	0.1421
10	-4914.8764	0.0000	-0.0098	-	-	-	0.5008	0.4992
11*	-4849.2302	0.9219	0.5750	0.4154	0.3124	0.2722	-	-
12	-4864.1171	0.0000	0.4356	0.5407	0.4115	-	0.0478	-
13	-4863.9848	0.0000	0.4390	0.5375	0.4099	-	-	0.0526
14	-4869.6133	0.0000	0.4153	0.5728	-	0.3750	0.0522	-
15	-4869.5024	0.0000	0.4190	0.5653	-	0.3756	-	0.0592
16	-4884.3557	0.0000	0.2768	0.8423	-	-	0.0743	0.0834
17	-4885.2463	0.0000	0.3518	-	0.5047	0.4327	0.0625	-
18	-4885.1041	0.0000	0.3570	-	0.4990	0.4281	-	0.0729
19	-4899.1158	0.0000	0.2207	-	0.7948	-	0.0959	0.1092
20	-4904.7257	0.0000	0.1960	-	-	0.7631	0.1080	0.1289
21	-4852.4506	0.0368	0.5912	0.3996	0.3012	0.2626	0.0366	-
22	-4852.3771	0.0396	0.5939	0.3991	0.2987	0.2606	-	0.0415
23	-4867.2128	0.0000	0.4551	0.5155	0.3882	-	0.0454	0.0509
24	-4872.6434	0.0000	0.4354	0.5412	-	0.3544	0.0488	0.0557
25	-4888.3092	0.0000	0.3736	-	0.4691	0.4026	0.0593	0.0690
26	-4855.5251	0.0017	0.6112	0.3851	0.2875	0.2531	0.0342	0.0401
bma_avg	-4849.4842	1.0000	0.5764	0.4141	0.3114	0.2713	0.0014	0.0017
highest	-4849.2302	0.9219	0.5750	0.4154	0.3124	0.2722	-	-
truth			0.6000	0.4000	0.3000	0.3000	0.0000	0.0000

Table 2: DGP is a convex combination,  $n = 2000$ ,  $\gamma_1 = 0.5$ , 1000 replications

Parameter	true	$W_1$	$W_2$	BMA	Convex
$\rho$	0.600	0.311	0.312	0.321	0.598
$\gamma_1$	0.5	-	-	-	0.499
intercept	1.000	1.733	1.738	1.713	1.006
$\beta_1$	2.000	2.033	2.032	2.030	2.000
$direct_1$	2.073	2.072	2.071	2.071	2.073
$indirect_1$	2.927	0.885	0.889	2.071	2.936
$total_1$	5.000	2.957	2.959	2.071	5.009
$\sigma^2$	1.000	1.168	1.163	1.153	1.001
post prob	-	0.443	0.557	-	-

Table 3: DGP is a convex combination,  $n = 2000$ ,  $\gamma_1 = 0.2$ , 1000 replications

	true	$W_1$	$W_2$	BMA	Convex
$\rho$	0.600	0.140	0.486	0.486	0.600
$\gamma_1$	0.200	-	-	-	0.196
intercept	1.000	2.159	1.292	1.292	1.000
$\beta_1$	2.000	2.101	2.004	2.004	2.000
<i>direct</i> <sub>1</sub>	2.112	2.109	2.113	2.113	2.114
<i>indirect</i> <sub>1</sub>	2.888	0.337	1.788	2.113	2.917
<i>total</i> <sub>1</sub>	5.000	2.446	3.901	2.113	5.031
$\sigma^2$	1.000	1.571	1.023	1.023	1.000
post prob	-	0.000	1.000	-	-

Table 4: DGP is a classic SAR (with  $W_1$  only),  $n = 2000$ , 1000 replications

	true	$W_1$	$W_2$	BMA	Convex
$\rho$	0.600	0.605	0.012	0.605	0.622
$\gamma_1$	1	-	-	-	0.973
intercept	1.000	0.986	2.509	0.986	0.944
$\beta_1$	2.000	1.998	2.184	1.998	1.998
<i>direct</i> <sub>1</sub>	2.194	2.198	2.185	2.198	2.198
<i>indirect</i> <sub>1</sub>	2.806	2.874	0.029	2.198	3.104
<i>total</i> <sub>1</sub>	5.000	5.072	2.214	2.198	5.302
$\sigma^2$	1.000	0.999	2.183	0.999	0.999
post prob	-	1.000	0.000	-	-



Table 5: Selection of the most relevant number of neighbors for each connectivity matrix.

# neighbors	Posterior probabilities (in %)			
	$W_{space}$	$W_{beds}$	$W_{baths}$	$W_{age}$
1	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0
12	0.0	<b>100.0</b>	0.0	0.0
13	0.0	0.0	<b>100.0</b>	0.0
14	<b>100.0</b>	0.0	0.0	<b>100.0</b>
15	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0

Table 6: Selection of the most relevant number of neighbors for each connectivity matrix using the Piribauer and Crespo Cuaresma (2016) approach

	$W_{space}$	$W_{beds}$	$W_{baths}$	$W_{age}$
# of neighbors	14	12	13	14
Probability (%) <sup>a</sup>	100.0	100.0	100.0	100.0

<sup>a</sup> The probability has been computed as the percentage of (retained) draws for which the number of neighbors in the above row has been preferred.

Table 7: Correlation of  $W_{space}$ ,  $W_{beds}$ ,  $W_{baths}$   $W_{age}$

	$W_{space}(14)$	$W_{beds}(12)$	$W_{baths}(13)$	$W_{age}(14)$
$W_{space}(14)$	1.000			
$W_{beds}(12)$	0.604	1.000		
$W_{baths}(13)$	0.642	0.490	1.000	
$W_{age}(14)$	0.728	0.511	0.538	1.000

Table 8: Posterior model probabilities using different convex combination of matrices

Model	log-marginal	Post. prob. (in %)	$\rho$	$W_{beds}$	$W_{baths}$	$W_{age}$
1	-25,216.558	0.0	0.795	0.412	0.588	0.000
2	-24,957.895	0.0	0.778	0.361	0.000	0.639
3	-24,172.776	0.0	0.791	0.000	0.462	0.538
4	-23,884.628	100.0	0.803	0.196	0.375	0.429
BMA	-23,884.628	-	0.803	0.196	0.375	0.429

Table 9: SAR model hedonic house price regressions

Estimates (variable)	$W_{space}$	$W_{beds}$	$W_{baths}$	$W_{age}$	$W_c = \gamma_1 W_{beds} + \gamma_2 W_{baths} + \gamma_3 W_{age}$
Constant	-0.358	-0.119	-0.052	-0.402	-0.513
	[-0.422 ; -0.297]	[-0.181 ; -0.053]	[-0.113 ; 0.005]	[-0.465 ; -0.342]	[-0.571 ; -0.453]
$\beta_1$ (ITLA)	0.391	0.351	0.314	0.384	0.306
	[0.383 ; 0.399]	[0.341 ; 0.360]	[0.304 ; 0.322]	[0.376 ; 0.392]	[0.298 ; 0.316]
$\beta_2$ (ILSIZE)	0.059	0.076	0.072	0.070	0.062
	[0.055 ; 0.063]	[0.072 ; 0.081]	[0.068 ; 0.076]	[0.066 ; 0.074]	[0.058 ; 0.066]
$\rho$ (Wy)	0.739	0.730	0.751	0.739	0.803
	[0.733 ; 0.746]	[0.723 ; 0.738]	[0.745 ; 0.759]	[0.733 ; 0.746]	[0.796 ; 0.811]
$\gamma_1$ ( $W_{beds}$ )	-	-	-	-	0.195
					[0.177 ; 0.218]
$\gamma_2$ ( $W_{baths}$ )	-	-	-	-	0.376
					[0.357 ; 0.397]
$\gamma_3$ ( $W_{age}$ )	-	-	-	-	0.429
					[0.407 ; 0.449]
Direct(ITLA)	0.422	0.381	0.342	0.415	0.330
	[0.413 ; 0.430]	[0.371 ; 0.391]	[0.332 ; 0.351]	[0.406 ; 0.423]	[0.321 ; 0.340]
Direct(ILSIZE)	0.064	0.083	0.078	0.076	0.067
	[0.059 ; 0.068]	[0.078 ; 0.088]	[0.074 ; 0.082]	[0.072 ; 0.081]	[0.063 ; 0.072]
Indirect(ITLA)	1.077	0.920	0.919	1.054	1.229
	[1.043 ; 1.116]	[0.885 ; 0.955]	[0.884 ; 0.950]	[1.021 ; 1.094]	[1.177 ; 1.292]
Indirect(ILSIZE)	0.163	0.200	0.210	0.193	0.249
	[0.151 ; 0.176]	[0.186 ; 0.213]	[0.197 ; 0.222]	[0.181 ; 0.206]	[0.230 ; 0.267]
Total(ITLA)	1.499	1.301	1.261	1.469	1.560
	[1.460 ; 1.542]	[1.261 ; 1.339]	[1.218 ; 1.299]	[1.430 ; 1.514]	[1.500 ; 1.629]
Total(ILSIZE)	0.226	0.283	0.289	0.269	0.316
	[0.210 ; 0.244]	[0.264 ; 0.301]	[0.271 ; 0.305]	[0.253 ; 0.286]	[0.293 ; 0.338]
Log-marginal likelihood	-26,490.735	-28,573.085	-26,903.300	-26,055.200	-23,884.628
Monte Carlo Errors					
Constant	0.0008	0.0007	0.0006	0.0007	0.0005
$\beta_1$ (ITLA)	0.0001	0.0001	0.0001	0.0001	0.0001
$\beta_2$ (ILSIZE)	0.0001	0.0001	0.0001	0.0001	0.0001
$\rho$ (Wy)	0.0001	0.0001	0.0001	0.0001	0.0001
$\gamma_1$ ( $W_{beds}$ )	-	-	-	-	0.0002
$\gamma_2$ ( $W_{baths}$ )	-	-	-	-	0.0002
$\gamma_3$ ( $W_{age}$ )	-	-	-	-	0.0002

Median values of the posterior distribution are reported. Further, figures between square brackets represent the credible confidence interval at the 99% threshold.

Table 10: Comparison of SAR models

Estimates (variable)	$W_{space}$	$W_c = \gamma_1 W_{beds} + \gamma_2 W_{baths} + \gamma_3 W_{age}$
Constant	0.475 [0.380 ; 0.567]	-0.513 [-0.571 ; -0.453]
$\beta_1$ (ITLA)	0.331 [0.320 ; 0.341]	0.306 [0.298 ; 0.316]
$\beta_2$ (ILSIZE)	0.056 [0.052 ; 0.060]	0.062 [0.058 ; 0.066]
Bedrooms (#)	0.009 [0.005 ; 0.014]	-
Bathrooms (#)	0.041 [0.036 ; 0.046]	-
House age	-0.009 [-0.011 ; -0.007]	-
$\rho$ (Wy)	0.702 [0.694 ; 0.709]	0.803 [0.796 ; 0.811]
$\gamma_1$ ( $W_{beds}$ )	-	0.195 [0.177 ; 0.218]
$\gamma_2$ ( $W_{baths}$ )	-	0.376 [0.357 ; 0.397]
$\gamma_3$ ( $W_{age}$ )	-	0.428 [0.407 ; 0.449]
Direct(ITLA)	0.352 [0.340 ; 0.363]	0.330 [0.321 ; 0.340]
Direct(ILSIZE)	0.060 [0.056 ; 0.064]	0.067 [0.063 ; 0.072]
Direct(Bedrooms (#))	0.010 [0.005 ; 0.015]	-
Direct(Bathrooms (#))	0.044 [0.038 ; 0.049]	-
Direct(House age)	-0.010 [-0.012 ; -0.007]	-
Indirect(ITLA)	0.757 [0.722 ; 0.791]	1.229 [1.177 ; 1.292]
Indirect(ILSIZE)	0.129 [0.119 ; 0.139]	0.249 [0.230 ; 0.267]
Indirect(Bedrooms (#))	0.021 [0.010 ; 0.032]	-
Indirect(Bathrooms (#))	0.094 [0.083 ; 0.106]	-
Indirect(House age)	-0.021 [-0.026 ; -0.015]	-
Total(ITLA)	1.109 [1.069 ; 1.151]	1.560 [1.500 ; 1.629]
Total(ILSIZE)	0.189 [0.175 ; 0.203]	0.316 [0.293 ; 0.338]
Total(Bedrooms (#))	0.032 [0.015 ; 0.047]	-
Total(Bathrooms (#))	0.137 [0.121 ; 0.155]	-
Total(House age)	-0.030 [-0.038 ; -0.022]	-
Log-marginal likelihood	-26052.668	-23884.639
Monte Carlo Errors		
Constant	0.0010	0.0005
$\beta_1$ (ITLA)	0.0001	0.0001
$\beta_2$ (ILSIZE)	0.0001	0.0001
Bedrooms (#)	0.0001	-
Bathrooms (#)	0.0001	-
House age	0.0001	-
$\rho$ (Wy)	0.0001	0.0001
$\gamma_1$ ( $W_{beds}$ )	-	0.0002
$\gamma_2$ ( $W_{baths}$ )	-	0.0002
$\gamma_3$ ( $W_{age}$ )	-	0.0002

Median values of the posterior distribution are reported. Further, figures between square brackets represent the credible confidence interval at the 99% threshold.