LEM
LILLE ÉCONOMIE MANAGEMENT

Document de travail du LEM / Discussion paper LEM
2019-05 “Version révisée”

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

Nicolas DEBARSY

LEM UMR 9221 / nicolas.debarsy@cnrs.fr

James P. LESAGE

Texas State University-San Marcos: Department of Finance & Economics

http://lem.cnrs.fr/

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-San Marcos
Department of Finance & Economics
601 University Drive
San Marcos, TX 78666, USA
jlesage@spatial-econometrics.com

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 econometrics 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. This type of model specification raises the question — which connectivity matrices should be used and which should be ignored. For example, in the case of L candidate weight matrices, there are $M = 2^L - L - 1$ possible ways to employ *two or more* of the L weight matrices in alternative model specifications. When $L = 5$, we have $M = 26$ possible models involving two or more weight matrices, and for $L = 10$, $M = 1,013$. We use 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. We focus on MCMC estimation for a set of M models, estimates of posterior model probabilities, model averaged estimates of the parameters, scalar summary measures of the non-linear partial derivative impacts, and associated empirical measures of dispersion for the impacts.

KEYWORDS: Markov Chain Monte Carlo estimation, SAR, Bayesian model averaging, 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 weight 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 Hazir et al. (2018). The convex combination approach uses a single $n \times n$ weight matrix $W_c(\Gamma) = \sum_{\ell}^{L} \gamma_\ell W_\ell$, with $0 \leq \gamma_\ell \leq 1,\ \ell = 1, \cdots, L$, $\sum_{\ell=1}^{L} \gamma_\ell = 1$ and $\Gamma = (\gamma_1, \cdots, \gamma_\ell, \cdots, \gamma_L)'$, constructed from alternative underlying types of connectivity between n observations reflected by $n \times n$ matrices W_ℓ. The resulting $W_c(\Gamma)$ can be used to specify dependence between n observations based on a convex combination of L different types of connectivity between observations. The scalar parameters γ_ℓ indicate the relative importance assigned to each type of dependence in the global cross-sectional dependence scheme. The two sets of constraints imposed reflect the fact that this approach relies on a convex combination. When each $W_\ell, \ell = 1, \cdots, L$, is row-normalized, then $W_c(\Gamma)$ obeys the conventional row-normalization.[1]

[1] This property is very useful as it avoids renormalizing W_c for each new set of values of $\gamma_\ell, \ell = 1, \cdots, L$ drawn in the estimation procedure.

This approach based on convex combinations avoids the issue of scaling for different metrics used in a generalized measure of distance by casting the problem as one of relative distance/proximity, inherent in conventional spatial regression weight matrices.

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 $L = 10$, $M = 1,013$. We use 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.

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.[2] A number of challenges for estimation and inference arise in this spatial autoregressive model where the connectivity scheme $W_c(\Gamma)$ is function of estimated parameters γ_ℓ. One is that the log-determinant term that arises in the likelihood function (and the conditional distribution for the spatial dependence parameters of the model) cannot be pre-calculated over a grid of values for the spatial dependence parameters, as is conventionally done in single weight matrix spatial regression models.[3] This is because changes in the parameters γ_ℓ lead to a new matrix $W_c(\Gamma)$ and associated log-determinant term $|I_n - \rho W_c(\Gamma)|$. A second issue relates to dealing with the restrictions that need to be imposed on the parameters γ_ℓ during estimation: $0 \leq \gamma_\ell \leq 1,\ \ell = 1, \cdots, L$ and $\sum_{\ell=1}^{L} \gamma_\ell = 1$. A third problem involves calculating simulated (empirical) measures of dispersion for the partial derivatives $\partial y/\partial x$ used to interpret estimates from this model.

We draw on earlier work by Pace and LeSage (2002) who use a Taylor series approximation for the log-determinant term in our model to address the first issue. A fourth-order Taylor series approximation to the log-determinant is set forth that uses traces calculated from products of the

[2] They focus on issues pertaining to proper use and interpretation of models involving convex combinations of connectivity matrices, but do not address the challenges regarding estimation and inference explored here, nor the Bayesian model averaging solution to account for model uncertainty.

[3] For spatial autoregressive processes, the interval $-1 < \rho < 1$ ensures a positive definite variance-covariance matrix.

multiple connectivity matrices prior to MCMC sampling. This allows rapid calculation of an approximation to the log-determinant term for any given set of γ_ℓ and ρ values during MCMC sampling.

The issue of restrictions on the parameters γ_ℓ, is addressed using a reversible jump MCMC block sampling scheme for initial draws of these parameters. A block of candidate values for γ_ℓ is proposed that obey the above mentioned restrictions and this block of parameters is then accepted or rejected in a Metropolis-Hastings (M-H) step. After some number of initial draws that quickly move the parameters towards their posterior means, a more conventional tuned random-walk procedure is used to produce proposal values.

The third challenge involves calculating (empirical) measures of dispersion for the partial derivatives $\partial y/\partial x$ that LeSage and Pace (2009) label *effects estimates*. A measure of dispersion for the (non-linear) effects is typically constructed by evaluating the partial derivatives using a large number (say 1,000) Markov Chain Monte Carlo (MCMC) draws for the parameters. Expressions for the partial derivatives require knowledge of the main diagonal elements of an $n \times n$ matrix inverse. For the case of single weight matrix, LeSage and Pace (2009) show how to use a (stochastic) trace approximation to find the main diagonal elements of the matrix inverse without calculating the full matrix inverse thousands of times. However, their approach does not immediately apply to the model developed here. We mix exact and approximate low-order traces based on the Taylor series approximation with stochastic estimates of higher-order traces to produce an extension of the method from LeSage and Pace (2009) that avoids calculation of the matrix inverse.

Having addressed issues associated with estimation of a spatial autoregressive model with a convex combination of connectivity matrices, there is still the question of which matrices are most consistent with the sample data and which are not. As noted, a number $M = 2^L - L - 1$ possible models arise for a set of L different matrices. We show how Metropolis-Hastings guided Monte Carlo integration of the joint posterior can be used during MCMC estimation of the models to produce log-marginal likelihoods and associated posterior model probabilities for each of the M possible models. This allows posterior inference regarding which connectivity matrices are most consistent with the sample data and also provides the basis for a Bayesian model averaged set of estimates.

In Section 2, we set forth matrix expressions for the model and discuss the MCMC estimation approach. 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 trace approximation to the log-determinant term that arises in conditional distributions of the model parameters; 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; an analysis of computational efficiency; and our approach to Bayesian model averaging.

Section 3 provides results from Monte Carlo experiments that explores bias and coverage to validate our estimation procedure. An applied illustration of the method in a hedonic house price model is the subject of section 4. Finally, section 5 concludes.

2 Computationally efficient expressions for the model

The spatial autoregressive (SAR) model that we wish to estimate is shown in (1), where each W_ℓ 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_ℓ reflect that observation j exhibits interaction with observation i, with different connectivity matrices describing different possible types of interaction (e.g., spatial, economic, and so on).

$$
\begin{aligned}
y &= \rho W_c(\Gamma) y + X\beta + \varepsilon \qquad (1)\\
W_c(\Gamma) &= \sum_{\ell=1}^{L} \gamma_\ell W_\ell \\
& \quad 0 \leq \gamma_\ell \leq 1 \\
& \quad \sum_{\ell=1}^{L} \gamma_\ell = 1
\end{aligned}
$$

The $n \times k$ matrix X in (1) contains explanatory variables, with β being the associated $k \times 1$ vector of parameters. The $n \times 1$ vector ε represents a constant variance normally distributed disturbance term, $\varepsilon \sim N(0, \sigma^2 I_n)$.

The SAR model in (1) can be expressed as shown in (2), a computationally convenient expression that isolates the parameters $\rho, \gamma_\ell, \ell = 1, \ldots, L$ in the $(L+1) \times 1$ vector ω.

$$
\begin{aligned}
\tilde{y}\omega &= X\beta + \varepsilon \qquad (2)\\
\tilde{y} &= (y,\ W_1 y,\ W_2 y,\ \ldots,\ W_L y)\\
\omega &= \begin{pmatrix} 1 \\ -\rho\gamma_1 \\ -\rho\gamma_2 \\ \vdots \\ -\rho\gamma_L \end{pmatrix} = \begin{pmatrix} 1 \\ -\rho\Gamma \end{pmatrix},\ \Gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_L \end{pmatrix}
\end{aligned}
$$

The value of isolating the parameter vector ω is that this allows us to pre-calculate the $n \times (L+1)$ matrix $\tilde{y}$ prior to the beginning of the MCMC sampling loop.

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

$$
\begin{aligned}
f(y|X, \mathcal{W}; \rho, \Gamma, \sigma^2, \beta) &= |R(\omega)| \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{e'e}{2\sigma^2}\right) \qquad (3)\\
e &= \tilde{y}\omega - X\beta\\
R(\omega) &= I_n - \rho W_c(\Gamma)\\
&= I_n - \rho(\gamma_1 W_1 + \gamma_2 W_2 + \ldots + \gamma_L W_L)
\end{aligned}
$$

where $|R(\omega)|$ is the Jacobian of the transformation, which in this case depends on the parameters (ρ, Γ). 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.[4] 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 ρ, Γ, β can be represented as $p(\rho) \sim U(-1, 1)$, $p(\Gamma) \sim U(0, 1)$ and $p(\beta) \propto 1$. The noise variance σ^2, is restricted to positive values. It is customary to consider a flat prior to its 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

[4] Many authors use $\rho \in (-\lambda_n^{-1}, 1)$, where λ_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 ρ, we avoid computing the minimum eigenvalue of $W_c(\Gamma)$, which depends on the value taken by Γ and can rely on the geometric expansion of $R(\rho, \Gamma)^{-1}$ to compute the impacts associated to the model.

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 ρ, Γ, reflect proper (bounded) probability distributions, while the priors on σ^2 and β are unbounded and improper since the integral over their parameter space (Ω_{σ^2} and Ω_β) is not finite. In the context of alternative models involving different weight matrices, if we rely on the same uniform priors for ρ, Γ and the parameters β, σ^2 are integrated out, the joint posterior distribution for the dependence parameters ρ, Γ are proper under relatively unrestrictive assumptions as shown in Hepple (1995a,b).

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 σ can be integrated out, leading to (6), where $G()$ denotes the Gamma function.

$$
\begin{aligned}
p(\rho, \Gamma, \sigma^2, \beta | y, X, \mathcal{W}) &\propto |R(\omega)|(\sigma^2)^{-\frac{(n+2)}{2}} \exp\left(-\frac{1}{2\sigma^2} e'e\right) && (4) \\
e &= \tilde{y}\omega - X\beta \\
p(\rho, \Gamma, \beta | y, X, \mathcal{W}) &\propto |R(\omega)| \int_0^\infty \sigma^{-(n+2)} \exp\left(-\frac{1}{2\sigma^2} e'e\right) d\sigma && (5) \\
&\propto |R(\omega)| \left(e'e\right)^{-n/2} && (6)
\end{aligned}
$$

To integrate out the k different β 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 ω shown in (7), with the term $|X'X|^{-1/2}$ and the exponent $-(n-k)/2$ arising from this integration (see Hepple, 1995a,b). This expression must be numerically integrated 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 that: (1) $n > k$, (2) $(X'X)^{-1}$ exists, and (3) $(y'MWy)^2 \neq y'W'MWyy'My$, Conditions (1) and (2) are not restrictive. Condition (3) essentially means that a valid error covariance matrix exists for the SAR model. This condition would be met in all cases where the single weight W is defined so that all observations are connected and a normalization is applied.[5] This result extends to our model immediately by recognizing that $W_c = \gamma_1 W_1 + \ldots + \gamma_L W_L$ is equivalent to the single weight matrix W, given our restrictions on the parameters $\gamma_\ell, \ell = 1, \ldots, L$, that $0 \leq \gamma_\ell < 1$, and $\sum_{\ell=1}^{L} \gamma_\ell = 1$. Hence, condition (3) should not be an important restriction to ensure posterior propriety of the joint distribution for the parameters ρ, Γ in our model.

2.2 The Markov Chain Monte Carlo (MCMC) estimation scheme

We note that 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_ℓ reflect important sample data in this type of model, as we wish to make distinctions between alternative specifications of W_ℓ. Highly correlated connectivity matrices will lead to problems identifying the parameters Γ. Distinguishing between alternative interaction structures also requires that spatial dependence reflected by the parameter ρ is different from zero, which should be clear when considering that for $\rho = 0$, the parameters Γ are not identifiable.

Conditional distributions for the model parameters required to implement MCMC estimation of the SAR specification in (1) are set forth here. Since our focus is on large samples n, reliance on the flat and uniform priors will not likely impact posterior estimates. The constraint $(-1 < \rho < 1)$ is imposed during MCMC estimation using rejection sampling. The interval $[0, 1]$ for γ_ℓ, $\ell = 1, \ldots, L$ is imposed during MCMC estimation, and we impose $\sum_{\ell=1}^{L} \gamma_\ell = 1$, by setting $\gamma_L = (1 - \sum_{\ell=1}^{L-1} \gamma_\ell)$. We discuss how proposal values for the vector of parameters Γ are generated later.

The conditional distribution for the parameters β is multivariate normal with mean and variance-covariance shown in (8).

[5]Multiple approaches to normalizing the matrix W have appeared in the spatial econometrics literature, all of which ensure that the maximum eigenvalue of the matrix W is one.

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

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

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

The (log) joint posterior for the parameters in ω after integrating out the parameters β and, σ^2 takes the form in (10). Details regarding a computationally efficient approach to calculating the approximation for the log-determinant term are postponed to the next section. Further, expression (10) does not reflect a known distribution (as in the case of the conditional distributions for β and σ^2).

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

We note that F consists of only sample data, so this expression can be calculated prior to MCMC sampling, leading to a computationally efficient expression reflecting a quadratic form: $\ln(\omega' F \omega)$, that can be easily evaluated for any vector of dependence parameters ω.

Recall $\omega' = (1,\ -\rho\gamma_1,\ -\rho\gamma_2,\ \ldots,\ -\rho\gamma_L)$, so we sample the parameter ρ from the joint posterior distribution in (10) conditioning on Γ and similarly for Γ conditioning on ρ. Details about the sampling procedure are postponed to section 2.4.

2.3 Log-determinants based on trace approximations

Pace and LeSage (2002) set forth a Taylor series approximation for the log-determinant of a matrix similar to our expression: $\ln|R(\rho, \Gamma)|$. They show that for a *symmetric* nonnegative connectivity matrix $W_c(\Gamma)$ with eigenvalues $\lambda_{\min} \geq -1, \lambda_{\max} \leq 1$, and $1/\lambda_{\min} < \rho < 1$, and $\mathrm{tr}(W_c(\Gamma)) = 0$ (where tr represents the trace):

$$\begin{aligned}
\ln(I_n - \rho W_c) &= -\sum_{i=1}^{\infty} \rho^i W_c^i / i \\
\ln|I_n - \rho W_c| &= -\sum_{i=1}^{\infty} \rho^i tr(W_c^i)/i \qquad (11) \\
&\simeq -\sum_{j=1}^{q} \rho^j tr(W_c^j)/j \\
tr(W_c) &= tr(\gamma_1 W_1 + \ldots + \gamma_L W_L) \\
&= \gamma_1 tr(W_1) + \ldots + \gamma_L tr(W_L) \qquad (12)
\end{aligned}$$

Golub and Van Loan (1996, pp. 566) provide the expression in (11), while (12) arises from linearity of the trace operator. Note that the 1st-order $tr(W_c)$ is zero, given the definitions of W_ℓ and the restrictions placed on γ_ℓ. The second-order trace can be expressed as a quadratic form involving the vector of parameters Γ and all pairwise multiplications of the individual matrices in W_ℓ as shown in (13) and (14), which are equivalent to expression (15).

$$\begin{aligned}
tr(W_c^2) &= \Gamma' \begin{pmatrix} tr(W_1^2) & tr(W_1 W_2) & \ldots & tr(W_1 W_L) \\ tr(W_2 W_1) & tr(W_2^2) & \ldots & tr(W_2 W_L) \\ \vdots & & \ddots & \\ tr(W_L W_1) & tr(W_L W_2) & \ldots & tr(W_L^2) \end{pmatrix} \Gamma, \qquad (13) \\
&= \Gamma' Q \Gamma. \qquad (14) \\
&= \sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j \, \mathrm{tr}(W_i W_j) \qquad (15)
\end{aligned}$$

LeSage and Pace (2009) point out that accelerated computation of traces can be accomplished using sums of matrix Hadamard products, shown in (16) for the case of *symmetric* matrices, and equal to $\sum_i^L \sum_j^L W_i \odot W_j'$. for *asymmetric* nonnegative connectivity matrices.

$$\begin{aligned} \mathrm{tr}(W_i W_j) &= \sum_k^N \sum_l^N w_{i,kl} w_{j,lk} \\ &= \sum_i^L \sum_j^L W_i \odot W_j \end{aligned} \tag{16}$$

Note that the formulation (14) separates the parameter vector Γ from the trace of the product of matrices, which allows pre-calculation of these traces prior to MCMC sampling. An even more efficient computational expression is $(\Gamma \otimes \Gamma)' vec(Q)$, where Q is the matrix composed of all the traces to be computed, $\otimes$ stands for the Kronecker product and $vec(Q)$ is the operator which stacks the columns of the matrix Q.

Using this approach leads to similar expressions for 3rd- and 4th-order traces, presented in (17) and (18) respectively.

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

$$\mathrm{tr}(W_c(\Gamma)^4) = \sum_{i=1}^L \sum_{j=1}^L \sum_{k=1}^L \sum_{l=1}^L \gamma_i \gamma_j \gamma_k \gamma_\ell \, \mathrm{tr}(W_i W_j W_k W_l) \tag{18}$$

Finally, the log-determinant term that arises in the conditional distributions from which we sample the spatial dependence parameters ρ and γ_ℓ, $\ell = 1, \ldots, L$ is approximated by a Taylor series approximation of order four, shown in (19).

$$\begin{aligned} \ln|R(\omega)| \simeq & -\rho^2 \, \mathrm{tr}(W_c(\Gamma)^2)/2 \\ & -\rho^3 (\mathrm{tr}(W_c(\Gamma)^3))/3 \\ & -\rho^4 (\mathrm{tr}(W_c(\Gamma)^4))/4 \end{aligned} \tag{19}$$

Section 3 reports results from Monte Carlo experiments that show this approximation of the log-determinant term works well to produce accurate parameter estimates for the model as well as good coverage intervals.

2.4 Sampling procedures for ρ and Γ

As noted in the introduction, a second computational challenge for MCMC estimation of the model is sampling parameters in the vector Γ, which must sum to one and cannot be negative. We set forth a reversible jump approach to block sampling Γ that proposes a vector of candidate values for $\gamma_\ell, \ell = 1, 2, \ldots, L-1$, with $\gamma_L = 1 - \sum_{\ell=1}^{L-1}$. The conditional distributions for the current and proposed vectors that we label Γ^c and Γ^p are evaluated with a M-H step used to either accept or reject the newly proposed vector Γ^p. Block sampling the parameter vector Γ 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 Γ.

The reversible jump procedure involves (for each $\gamma_\ell, \ell = 1, \ldots, L-1$) a three-headed coin flip. By this we mean a uniform random number on the closed interval *coin flip* $= U(0,1)$, with head #1 a value of *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 γ_ℓ^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 $(\gamma_\ell^c, 1]$.

Green (1995) proposed the reversible jump MCMC method as a generalization of M-H for the purpose of sampling from conditional posterior distributions. In our case, we make a random choice between a new vector that increments some values of γ_ℓ and decrements others. In theory we can produce samples from the correct posterior using just *increment* and *decrement* steps, but adding the third step that leaves a value of γ_ℓ unchanged (which we label a *stay* step) was found to improve the acceptance rate of the sampler. This finding is consistent with results in Richardson and Green (1997) and Dennison et al. (2002). Given equal probabilities and a parameter space that is discrete as opposed to continuous, the reversible jump procedure from Green (1995) simplifies considerably.

Green (1995) sets forth a general form of acceptance probability shown in (20), where $p(\Gamma^p|\cdot)$ is our conditional distribution, $\pi(\Gamma)$ is the prior, $Q(\Gamma^c|\Gamma^p)$ is the probability of proposing a decrement (increment, stay), and $|J|$ is a Jacobian term to account for the change in scale that arises if the parameter dimension changes. Since the prior probabilities are equal, $\pi(\Gamma^c)/\pi(\Gamma^p) = 1$, the parameter dimension is constant so $|J| = 1$, and we use equal probabilities of proposing increments/decrements/stays, the M-H acceptance probability simplifies to the expression shown in (21). The (non-logged) conditional distributions in expression (22) are used to calculate a M-H acceptance probability, where we use $\cdot$ to denote the conditioning parameter (ρ).

$$\psi_{MH}(\Gamma^c, \Gamma^p) = \min\{1, \frac{p(\Gamma^p|\cdot)}{p(\Gamma^c|\cdot)} \frac{\pi(\Gamma^p)}{\pi(\Gamma^c)} \frac{q(\Gamma^c|\Gamma^p)}{q(\Gamma^p|\Gamma^c)} |J|\} \tag{20}$$

$$= \min\{1, \frac{p(\Gamma^p|\cdot)}{p(\Gamma^c|\cdot)}\} \tag{21}$$

$$= \min\left(1, \exp[(ln\ p(\Gamma^p|\cdot) - ln\ p(\Gamma^c|\cdot)]\right) \tag{22}$$

The expression to be evaluated at the current and proposed vectors of parameters Γ 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 Γ. 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 Γ using the pre-calculated traces and the conditioning parameter ρ. From the expression in (19), it should be clear that we 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 $\tilde{y} = (y\ W_1 y\ W_2 y\ \ldots\ W_L y)$ and the vector $\omega' = (1,\ -\rho\gamma_1,\ -\rho\gamma_2,\ \ldots,\ -\rho\gamma_L)$.

After the initial 1,000 draws we switch to a tuned random-walk proposal procedure. This is needed because proposals from the reversible jump 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 $j = 1, \ldots, L$ 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 Γ. This is used in conjunction with the standard deviations and the block sampling procedure described previously to produce proposals. As already noted, we carry out a series of Monte Carlo experiments in Section 3 that allow us to assess the efficacy of this approach to sampling the parameters γ_ℓ.

Finally, the parameter ρ is drawn from the joint distribution for ω shown in (10), conditioning on Γ. 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 ρ. 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 $\rho, \beta, \Gamma, \sigma^2$, we need to compute the reduced form of the model to compute the partial derivatives of y with respect to the relevant exogenous determinant to interpret its effect. For the standard SAR model, LeSage and Pace (2009) derive these partial derivatives and further develop scalar summary measures which represent own- and cross-partial derivatives that they label *average direct total* and *indirect* effects. For the SAR model with a convex combination of connectivity matrices, the partial derivative of the outcome variable with respect to the r^{th} determinant is shown in (23) while the associated scalar summary measures are presented in (24) to (26), where ι_n is an $n \times 1$ vector of ones.

$$
\begin{aligned}
\partial y/\partial x^r &= = S_r(W_c(\Gamma)) && (23)\\
S_r(W_c(\Gamma)) &= (I_n - \rho W_c(\Gamma))^{-1}\beta_r \\
&= I_n\beta_r + \rho W_c(\Gamma)\beta_r + \rho^2 W_c(\Gamma)^2\beta_r + \ldots \\
\bar{M}(r)_{direct} &= n^{-1} tr(S_r(W_c(\Gamma))) && (24)\\
\bar{M}(r)_{total} &= n^{-1}\iota_n' S_r(W_c(\Gamma))\iota_n && (25)\\
\bar{M}(r)_{indirect} &= \bar{M}(r)_{total} - \bar{M}(r)_{direct} && (26)\\
W_c(\Gamma) &= \sum_{\ell=1}^{L} \gamma_\ell W_\ell
\end{aligned}
$$

While expressions in (24), (25) and (26) produce point estimates for the scalar summary measures of effects, we also require measures of dispersion for the purpose of statistical tests regarding the significance of these effects. Use of an empirical distribution constructed by simulating the non-linear expressions in (23) using (say 1,000) draws from the posterior distribution of the underlying parameters $\rho, \beta_r, \gamma_\ell, \ell = 1, \ldots, L$ is suggested by LeSage and Pace (2009).

Note that a naive approach to such a simulation-based empirical distribution would require calculation of the $n \times n$ matrix inverse $S_r(W_c(\Gamma))$ a large number of times, for varying values of the parameters $\rho, \beta_r, \gamma_\ell, \ell = 1, \ldots, L$, which would be computationally intensive. For the traditional SAR model, LeSage and Pace (2009) show that the required quantity for constructing the empirical distribution of the effects is $\text{tr}(S_r(W))$, which can be estimated without a great deal of computational effort. For the purpose of calculating the effects estimates, LeSage and Pace (2009) set forth a procedure that relies on a $(1 \times (q+1))$ vector R, shown in (27), containing average diagonal elements from powers of W, and the $(1 \times (q+1))$ vector g in (28) and corresponding diagonal matrix G shown in (29), and $(q+1) \times 1$ vector of ones ι_{q+1}.[6]

$$
\begin{aligned}
R &= \begin{pmatrix} 1 & 0 & \text{tr}(W^2)/N & \text{tr}(W^3)/N & \dots & \text{tr}(W^q)/N \end{pmatrix} &(27)\\
g &= \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^q \end{pmatrix} &(28)\\
G &= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \rho & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \rho^q \end{pmatrix} &(29)\\
\bar{M}(r)_{direct} &= \beta_r R G \iota_{q+1} &(30)\\
\bar{M}(r)_{total} &= \beta_r g \iota_{q+1} &(31)\\
\bar{M}(r)_{indirect} &= \bar{M}(r)_{total} - \bar{M}(r)_{direct} &(32)
\end{aligned}
$$

Given the (pre-calculated) traces, empirical measures of dispersion for the effects can be constructed using MCMC draws for the parameter ρ in g, G and β_r in expressions (30), (31), where we note that the total effects are the sum of the direct plus indirect effects.

In practice, computational implementations do not actually calculate traces but rather rely on estimates of these which do not require much computational effort. Specifically, an iterative procedure is used to produce a Monte Carlo estimate of the traces. A set of $j = 100$ iterations shown in (33) involving $m = 150$ vectors of *iid* normal deviates v are averaged as shown in (34), to produce Monte Carlo estimates of the diagonals, (see Barry and Pace, 1999).

$$
\begin{aligned}
v_{(j)} &= W v_{(j-1)} &(33)\\
\text{tr}(W^j) &\simeq (v \odot v_{(j)}) \frac{\iota_m}{m} &(34)
\end{aligned}
$$

Our situation differs because the matrix $W_c(\Gamma)$ depends on estimated parameters $\gamma_\ell, \ell = 1, \dots, L$ ruling out use of pre-calculated traces. One could

[6] Computational implementations of this approach in the *Spatial Econometrics Toolbox* and the $R-$ language *Spdep* package set $q = 100$.

rely on posterior means for Γ, labeled $\overline{\Gamma}$ to create a single matrix $W_c(\overline{\Gamma})$, for which pre-calculated traces could be used in (27). However, this would ignore stochastic variation in the effects estimates that arise from the fact that there is uncertainty regarding the parameter vector Γ. Ideally, we would like to use draws for the Γ parameters from their posterior distributions when simulating the empirical distribution of effects estimates.

We have already calculated the first four order traces to produce the Taylor series approximation to the log-determinant term. Our approach for efficiently estimating the empirical distribution for partial derivative effects required for inference involves the following steps. We calculate $q = 100$ estimated traces using $W_c(\overline{\Gamma}) = \sum_{\ell=1}^{L} \bar{\gamma}_\ell W_\ell$ in place of W in expression (34). The second- through fourth-order (estimated) traces are replaced with those shown in (15), (17) and (18) during simulation, where MCMC draws for the parameters γ_ℓ are used instead of relying on the posterior mean value of these parameters. Note that this incorporates uncertainty in the parameters γ_ℓ for low-order traces by using MCMC draws for these parameters. Given that the vector products in (30) and (31) involve increasingly small magnitudes associated with higher-order powers of the parameters ρ and Γ, 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 ρ (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. The Monte Carlo experiments carried out in Section 3 will be used to assess the accuracy of our approximation approach to constructing the effects estimates and the associated inference.

2.6 Log marginal likelihoods

Calculation of the log marginal likelihood for the convex combination of weight matrices model would involve integrating the joint posterior distribution over all model parameters. We can analytically integrate out the parameters β, and σ^2, leading to a (log kernel) joint posterior for the remaining model parameters in ω that takes the form in (10).

Integrating out the remaining parameters using numerical methods could be computationally intensive for cases involving a large number of parameters in the vector Γ. Monte Carlo numerical integration would tackle the problem by sampling a large number of parameters in ω, but this would be inefficient as many of the samples would reflect areas of low support in the joint posterior distribution. However, we have a set of draws for the parameters ω based on Metropolis-Hastings that tune the parameters in the vector ω to areas of high density in the joint posterior. Evaluating the joint posterior on every trip through the MCMC sampling loop for the values in the vector ω, allows us to take the mean of these evaluations to produce a Metropolis-Hastings tuned Monte Carlo estimate of the (kernel) joint posterior.

We can then add the constants to produce an estimate of the log marginal likelihood. The constants take the form: $\kappa = -log(1/\max\rho - 1/\min\rho) + log(Gamma(dof)) - dof \times log(2\pi)$, where $log(Gamma())$ 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 $exp[log\ p(\omega) - \max(log\ p(\omega))]$.

To illustrate these ideas, a model was generated based on $n = 1,500$ and three connectivity matrices, W_1, a two-nearest neighbors connectivity matrix, W_2 a three nearest neighbors connectivity matrix, both based on the *same* set of random latitude-longitude coordinates, and W_3 a 12 nearest neighbors connectivity matrix based on a *different* random set of latitude-longitude coordinates. This results in the matrices W_1, W_2 being highly correlated. The two sets of coordinates have been drawn from standard normal distributions. We also set $\gamma_1 = 0.4, \gamma_2 = 0.2, \gamma_3 = 0.4$ and $\rho = 0.6$. We consider $X = [\iota_n, X_1, X_2]$, where X_1 and X_2 come from standard normal distributions, $\beta = [-1, -0.5, 1.5]$, and an error term also drawn from a standard normal distribution.

Figure 1 shows the normalized non-logged joint posterior points evaluated during MH-MC integration for the parameters γ_1, γ_2, which are centered on the true values of $\gamma_1 = 0.4, \gamma_2 = 0.2$, despite the high correlation between the W_1, W_2 matrices.[7] Use of redundant information in components of the convex combination of connectivity matrices will produce correlation in the associated parameters γ_1, γ_2. We might expect to see a high negative correlation because two highly correlated weight matrices act as substitutes for each other in the model. Higher values of γ_1 that emphasize the role of W_1 are associated with lower values of γ_2 since emphasis on W_2 would result in redundant information regarding the spatial dependence structure. It is also the case that the summing up constraint $\sum_{\ell=1}^{3} \gamma_\ell = 1$, produces a situation where higher values of γ_1 that emphasize the role of W_1 are associated

[7] LeSage and Pace (2014) suggest measuring correlation between spatial weight matrices using a standard random normal vector u, and constructing vectors W_1u, W_2u, for which a correlation coefficient can be calculated. This approach produced a correlation coefficient between W_1uW_2u equal to 0.8180, whereas that between W_1u and W_3u was 0.0516.

with lower values of γ_2. In fact, if W_1, W_2 were perfectly correlated, values of $\gamma_1 = 0.6$ and $\gamma_2 = 0$ would produce the same log-posterior as values of $\gamma_1 = 0$ and $\gamma_2 = 0.6$.

Figure 2 shows the joint posterior points evaluated for the parameters γ_1, γ_3, which is also centered on the true values of $\gamma_1 = 0.4, \gamma_3 = 0.4$. Here we see no discernible correlation between the posterior distribution for values of γ_1, γ_3 as the matrices W_1, W_3 reflect independent information.

Figure 3 shows a similar plot of points evaluated by our MH-MC integration scheme for the joint posterior distribution of the parameters ρ and γ_1. By definition, our model implies correlation between the parameter ρ and values taken by the parameters $\gamma_\ell, \ell = 1, \ldots, L$, as $\omega' = [1,\ -\rho\gamma_1,\ -\rho\gamma_2,\ -\rho(1-\gamma_1-\gamma_2)]$. We see negative covariance between ρ and γ_1, and in Figure 4 positive covariance between ρ, γ_3.

2.7 Computational speed

To provide an indication of the computation efficiency of the approach set forth, Table 1 shows the time required to produce model estimates.[8] This includes time required to: 1) pre-calculate the terms used in the fourth-order trace approximation, 2) to produce 20,000, 50,000 and 100,000 MCMC draws, and 3) to calculate effects estimates and the empirical distribution on which posterior inference would be based.

The results are for a data-generated sample based on three connectivity matrices each constructed using three independent sets of $n \times 1$ (standard) random normal vectors of latitude-longitude coordinates. The three connectivity matrices were based on 5, 8 and 10 nearest neighbors, and values of $\gamma_1 = 0.2, \gamma_2 = 0.5, \gamma_3 = 0.3$ were used.

The data generating process for the $n \times 1$ dependent variable vector is: $y = (I_n - \rho W_c(\Gamma))^{-1}\left[\sum_{k=1}^{3} X_k\beta_k + \varepsilon\right]$, where $X_k \sim N(0,1), k = 1,2,3$, and the $k \times 1$ vector $\beta = (1,\ 0,\ -1)'$, $\varepsilon \sim N(0,1)$, and $W_c(\Gamma) = (0.2W_1 + 0.5W_2 + 0.3W_3)$. Note that since W_1, W_2, W_3 are row-normalized nearest neighbor connectivity matrices, we require the asymmetric connectivity matrix approach to calculating traces.

The second portion of Table 1 normalizes time required to produce model estimates relative to the case of $N = 1,000$ and 20,000 MCMC draws. From this set of results, we can divide the last column by the first column to find that a five-fold increase in the number of draws leads to a four-fold increase

[8]Times reported are for a Dell XPS-15 laptop with Intel i9-8950HK 2.90 GHz CPU and Matlab 2018b.

Table 1: Performance results for 3 $W-$matrices

	Timing in seconds		
(Nobs/Draws)	20,000	50,000	100,000
1,000	3.26	6.82	13.34
5,000	10.43	22.44	41.31
10,000	19.67	40.85	76.93
25,000	35.74	62.39	107.76
	Times normalized		
1,000	1.00	2.08	4.08
5,000	3.19	6.87	12.65
10,000	6.02	12.51	23.55
25,000	10.94	19.10	32.99

in time required for $N = 1,000$, with (relative) times required declining to three for $N = 25,000$.

Considering the last row of the relative times divided by the first row shows the impact of increasing observations 25-fold on time needed to produce estimates. Here we see that for the case of 20,000 draws, the time increased by 10.9 times in the face of a 25-times increase in sample size N, and for 50,000 draws time increased by a factor of 9.1 = 19.1/2.08, and for 100,000 draws by 8.08=32.99/4.08. These results suggest that the estimation procedure scales very favorably in terms of both the number of MCMC draws and the number of observations N.

2.8 Bayesian model averaging

Given the speed of estimation for single models and the availability of multicore computer architecture, it is possible to estimate models based on all possible combinations of two or more spatial connectivity matrices, even in cases of 10 connectivity 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 β, Γ, ρ 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 in Table 2 the estimation results for all models involving two or more connectivity matrices using a set of five candidate $W-$matrices and a sample of $N = 2,000$ observations. The DGP used is shown in (35):

$$y = (I_n - \rho W_c(\Gamma))^{-1}(X\beta + \varepsilon) \tag{35}$$

where X includes a constant term and 2 standard normal variables, $\beta = (-1,\ -0.5,\ 1.5)$ and $\rho = 0.6$. The error term ε is assumed Normally distributed, centered around zero and with a variance $\sigma^2 I_n$ with $\sigma^2 = 3.6061$ so that the signal to noise ratio (SNR) of the model is equal to 0.7. In this paper, we define the SNR following Debarsy and LeSage (2018). Letting $A = (I_n - \rho W_c(\Gamma))^{-1} X\beta$, $X = [\iota_n,\ x_1]$, $\beta = [\beta_0, \beta_1]'$ and $B = (I_n - \rho W_c(\Gamma))^{-1}$, the SNR is defined as follows:

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

The five candidate $W-$matrices are all 5 nearest neighbors matrices but constructed from independent sets of random normal latitude-longitude vectors to ensure they convey different information content. The true values for $\gamma_1 = 0.4, \gamma_2 = 0.3, \gamma_3 = 0.3$ were used with $\gamma_4 = \gamma_5 = 0$. 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 for inference.

Estimates for ρ and the parameters Γ for the 26 possible models involving combinations of 2 or more of the five $W-$matrices are shown in the upper part of Table 2, 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 2 also shows posterior means for a set of model averaged estimates for the parameters $\rho, \gamma_i, i = 1, \ldots, 5$. 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 $\gamma_4 = 0.0014, \gamma_5 = 0.0017$. We can calculate credible intervals for these two parameters using the distribution of set of 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

Table 2: Estimates for 5 candidate $W-$matrices, $M = 26$ models, $N = 2000$.

Models	log marginal	Prob(m_i)	ρ	γ_1	γ_2	γ_3	γ_4	γ_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

Γ. That is, we can use credible intervals for the BMA posterior estimates of γ_4, γ_5 to see if an inference of zero for these parameters is reasonable. The 0.01 interval for γ_4 is 0.0046, and that for γ_5 is 0.005, allowing us to conclude that these parameters are sufficiently close to zero to make an inference of zero reasonable.

3 Monte Carlo experiments

To assess the validity of our MCMC estimation scheme a Monte Carlo study was carried out. This involved 500 experiments with varying sets of true parameter values used to generate sample data vectors y. A set of three standard normal explanatory variables were used with associated parameters ($\beta_1 = 1.0, \beta_2 = 0.0, \beta_3 = -1.0$), with the matrix X fixed in repeated sampling. Noise variances of $\sigma^2 = 1$ and $\sigma^2 = 10$ were used in conjunction with three values of ρ, equal to $(-0.6, 0.2, 0.6)$. The motivation for a value of $\rho = 0.2$ is that these models should perform poorly in the face of weak spatial dependence. Sample sizes of 500, 1,000 and 5,000 were used in an effort to see if expected asymptotic properties hold true, with bias and mean-squared error (MSE) shrinking as sample size increases.

A set of 30,000 MCMC draws were used with the first 10,000 discarded for burn-in. In addition to analyzing estimates for the parameters $\rho, \Gamma, \beta, \sigma^2$, estimates of the direct and indirect effects associated with the three explanatory variables were also assessed.

Three connectivity matrices were used, based on two, four and six nearest neighbors, where each connectivity matrix was generated using different sets of standard normal latitude-longitude coordinates. This produces matrices that are reasonably uncorrelated. Values for the parameters $\gamma_1 = 0.5, \gamma_2 = 0.4, \gamma_3 = 0.1$ were used, with the value of 0.1 included because this should be more difficult to accuracy estimate.

Coverage estimates were also produced to see if the 95% credible intervals contained the true parameters. The percent of times (from the 500 trials) when the true values were within the 95% intervals was calculated.

We begin with a presentation of the 95% coverage results in Table 3. From the table, we see good coverage of the true parameter values for all of the 18 different Monte Carlo scenarios, including the cases where $n = 500, \rho = 0.2$ and the noise variance was relatively larger based on $\sigma^2 = 10$.

Table 4 presents bias results for the 18 Monte Carlo scenarios in the same format as Table 3. Here we see large bias for the $\gamma_3 = 0.1$ parameter in situations involving the smaller $N = 500$ sample size, the low $\rho = 0.2$

Table 3: Coverage results for 500 trials

	$N=500$			$N=1,000$			$N=5,000$		
$\sigma^2=1$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$
Parameters	95%	95%	95%	95%	95%	95%	95%	95%	95%
ρ	0.9660	0.9560	0.9620	0.9520	0.9420	0.9520	0.9420	0.9660	0.9480
β_1	0.9580	0.9420	0.9460	0.9340	0.9540	0.9420	0.9460	0.9420	0.9240
β_2	0.9500	0.9520	0.9400	0.9660	0.9640	0.9600	0.9680	0.9580	0.9500
β_3	0.9500	0.9380	0.9460	0.9580	0.9420	0.9540	0.9500	0.9620	0.9480
γ_1	0.9640	0.9620	0.9680	0.9740	0.9540	0.9540	0.9480	0.9500	0.9680
γ_2	0.9460	0.9600	0.9540	0.9560	0.9420	0.9420	0.9480	0.9360	0.9540
γ_3	0.9860	0.9760	0.9820	0.9700	0.9760	0.9720	0.9660	0.9860	0.9520
$direct_1$	0.9580	0.9400	0.9400	0.9360	0.9540	0.9420	0.9440	0.9420	0.9240
$direct_2$	0.9500	0.9520	0.9400	0.9660	0.9640	0.9600	0.9680	0.9580	0.9500
$direct_3$	0.9440	0.9400	0.9500	0.9620	0.9400	0.9580	0.9460	0.9580	0.9420
$indirect_1$	0.9700	0.9580	0.9660	0.9540	0.9540	0.9520	0.9400	0.9640	0.9520
$indirect_2$	0.9500	0.9640	0.9400	0.9660	0.9640	0.9600	0.9680	0.9580	0.9500
$indirect_3$	0.9520	0.9600	0.9700	0.9540	0.9500	0.9480	0.9440	0.9620	0.9420
$total_1$	0.9400	0.9520	0.9680	0.9440	0.9580	0.9480	0.9500	0.9580	0.9540
$total_2$	0.9500	0.9520	0.9400	0.9660	0.9640	0.9600	0.9680	0.9580	0.9500
$total_3$	0.9680	0.9500	0.9660	0.9560	0.9400	0.9520	0.9520	0.9540	0.9420
σ^2	0.9480	0.9620	0.9400	0.9520	0.9640	0.9600	0.9520	0.9540	0.9380
	$N=500$			$N=1,000$			$N=5,000$		
$\sigma^2=10$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$	$\rho=-0.6$	$\rho=0.2$	$\rho=0.6$
Parameters	95%	95%	95%	95%	95%	95%	95%	95%	95%
ρ	0.9500	0.9460	0.9640	0.9660	0.9640	0.9620	0.9420	0.9500	0.9620
β_1	0.9440	0.9600	0.9260	0.9480	0.9760	0.9520	0.9620	0.9480	0.9300
β_2	0.9520	0.9560	0.9480	0.9620	0.9560	0.9640	0.9460	0.9540	0.9460
β_3	0.9720	0.9580	0.9560	0.9520	0.9480	0.9580	0.9420	0.9440	0.9500
γ_1	0.9440	0.9740	0.9680	0.9560	0.9520	0.9640	0.9580	0.9460	0.9580
γ_2	0.9400	0.9860	0.9560	0.9380	0.9820	0.9440	0.9520	0.9560	0.9600
γ_3	0.9740	0.9840	0.9740	0.9780	0.9820	0.9720	0.9680	0.9660	0.9620
$direct_1$	0.9460	0.9600	0.9300	0.9440	0.9740	0.9440	0.9640	0.9480	0.9280
$direct_2$	0.9520	0.9560	0.9480	0.9620	0.9560	0.9640	0.9460	0.9540	0.9460
$direct_3$	0.9760	0.9540	0.9580	0.9520	0.9500	0.9600	0.9400	0.9420	0.9440
$indirect_1$	0.9560	0.9460	0.9780	0.9560	0.9580	0.9660	0.9540	0.9560	0.9600
$indirect_2$	0.9520	0.9720	0.9480	0.9620	0.9640	0.9640	0.9460	0.9540	0.9460
$indirect_3$	0.9620	0.9420	0.9580	0.9560	0.9500	0.9620	0.9480	0.9440	0.9600
$total_1$	0.9500	0.9540	0.9680	0.9480	0.9500	0.9720	0.9400	0.9580	0.9640
$total_2$	0.9520	0.9560	0.9480	0.9620	0.9560	0.9640	0.9460	0.9540	0.9460
$total_3$	0.9740	0.9520	0.9600	0.9560	0.9540	0.9640	0.9480	0.9440	0.9560
σ^2	0.9400	0.9660	0.9600	0.9460	0.9360	0.9620	0.9560	0.9440	0.9560

value of the spatial dependence parameter and the larger noise variance of

$\sigma^2 = 10$.

From the first part of the table showing results for the case where $\sigma^2 = 1$, we see bias for γ_3 of $(0.0257, 0.1383, 0.0178)$ for $N = 500$ and $\rho = (-0.6, 0.2, 0.6)$ respectively, which decreases to bias of $(0.0122, 0.0866, 0.0073)$ for $N = 1,000$ and further decreases to bias of $(-0.0017, 0.0245, -0.0034)$ for $N = 5,000$. As we would expect, the corresponding biases for the γ_3 parameter are larger for the same scenarios when the noise variance was increased to $\sigma^2 = 10$, specifically: $(0.0536, 0.1722, 0.0383)$ for $N = 500$, which decreases to bias of $(0.0247, 0.1233, 0.0216)$ for $N = 1,000$ and further decreases to bias of $(0.0032, 0.0427, -0.0017)$ for $N = 5,000$.

Another result from Table 4 is that bias for the indirect and total effects estimates are largest for values of the spatial dependence parameter $\rho = 0.6$, which of course produce the largest indirect and total effects magnitudes. For example, the true total effects magnitudes for X_1, X_2, X_3 equal 2.5, 0 and -2.5, respectively, for all scenarios This means that bias magnitudes of 0.18 and -0.18 in the case of $\sigma^2 = 1, N = 500$ represent 7 percent bias, and in the worse case scenario where $\sigma^2 = 10, N = 500$ the bias of 0.34 and -0.34 is around 14 percent.

Summarizing our Monte Carlo results regarding bias, it may be difficult to draw accurate inferences regarding models involving convex combinations of connectivity matrices in cases where the sample size is small (e.g., $N = 500$) and some of the connectivity matrices are relatively unimportant, e.g., $\gamma = 0.1$. It is important to note that accuracy of inferences would also depend on: signal-to-noise ratios, the strength of spatial dependence, as well as the amount of correlation between connectivity matrices considered.

Table 5 presents mean-squared error (MSE) results for the 18 Monte Carlo scenarios in the same format as the previous tables. Here we see the largest MSE for the indirect and total effects parameters, which are non-linear functions of the underlying parameters ρ, Γ, β. Errors made in estimates of these parameters becomes magnified by the matrix inverse $(I_n - \rho W_c(\Gamma))^{-1}$, which should be clear from the series expansion of this inverse: $(I_n - \rho W_c(\Gamma))^{-1} = I_n + \rho W_c(\Gamma) + \rho^2 W_c(\Gamma)^2 + \rho^3 W_c(\Gamma)^3 + \ldots$, where the worse case scenarios involve the smaller sample size of $N = 500$, with MSE decreasing as sample size increases (ceteris paribus). We should also keep in mind the larger magnitudes of the indirect and total effects scalar summaries for cases where $\rho = 0.6$.

Table 4: Bias results for 500 trials

	$N = 500$			$N = 1,000$			$N = 5,000$		
$\sigma^2 = 1$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$
Parameters	bias	bias	bias	bias	bias	bias	bias	bias	bias
ρ	-0.0068	0.0080	0.0091	-0.0062	0.0027	0.0010	0.0034	0.0022	0.0012
β_1	0.0006	-0.0018	-0.0010	-0.0013	-0.0034	-0.0004	0.0006	0.0000	-0.0005
β_2	-0.0020	0.0016	0.0032	0.0023	0.0004	-0.0026	-0.0003	-0.0002	-0.0003
β_3	-0.0011	-0.0026	-0.0014	0.0019	0.0022	-0.0017	0.0003	0.0002	0.0010
γ_1	-0.0061	-0.0677	-0.0069	-0.0049	-0.0293	-0.0005	0.0018	-0.0041	0.0017
γ_2	-0.0196	-0.0706	-0.0109	-0.0073	-0.0572	-0.0068	-0.0002	-0.0205	0.0017
γ_3	0.0257	0.1383	0.0178	0.0122	0.0866	0.0073	-0.0017	0.0245	-0.0034
$direct_1$	0.0011	-0.0011	0.0002	-0.0004	-0.0031	0.0001	0.0004	0.0001	0.0001
$direct_2$	-0.0021	0.0016	0.0034	0.0023	0.0004	-0.0027	-0.0003	-0.0002	-0.0003
$direct_3$	-0.0017	-0.0033	-0.0027	0.0010	0.0019	-0.0023	0.0005	0.0001	0.0005
$indirect_1$	-0.0010	0.0293	0.1786	-0.0015	0.0111	0.0583	0.0016	0.0052	0.0198
$indirect_2$	0.0009	0.0007	0.0071	-0.0009	0.0000	-0.0038	0.0001	-0.0001	-0.0003
$indirect_3$	0.0013	-0.0297	-0.1827	0.0013	-0.0115	-0.0601	-0.0020	-0.0052	-0.0193
$total_1$	0.0001	0.0282	0.1789	-0.0019	0.0080	0.0584	0.0020	0.0053	0.0199
$total_2$	-0.0011	0.0023	0.0104	0.0014	0.0004	-0.0065	-0.0002	-0.0003	-0.0006
$total_3$	-0.0003	-0.0330	-0.1854	0.0023	-0.0096	-0.0624	-0.0014	-0.0050	-0.0189
σ^2	0.0017	0.0022	0.0015	0.0022	0.0020	-0.0009	0.0002	0.0021	-0.0020
	$N = 500$			$N = 1,000$			$N = 5,000$		
$\sigma^2 = 10$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$
Parameters	bias	bias	bias	bias	bias	bias	bias	bias	bias
ρ	-0.0087	0.0049	0.0114	-0.0122	0.0077	0.0051	0.0009	0.0058	0.0020
β_1	0.0012	-0.0101	0.0012	0.0014	-0.0030	-0.0062	0.0002	-0.0009	0.0027
β_2	0.0058	-0.0028	-0.0098	0.0104	0.0051	0.0021	0.0010	-0.0000	-0.0003
β_3	-0.0021	-0.0114	-0.0003	-0.0057	0.0028	0.0046	0.0012	0.0012	0.0016
γ_1	-0.0176	-0.1072	-0.0160	-0.0074	-0.0651	-0.0014	-0.0003	-0.0143	0.0011
γ_2	-0.0360	-0.0650	-0.0223	-0.0173	-0.0582	-0.0202	-0.0029	-0.0285	0.0006
γ_3	0.0536	0.1722	0.0383	0.0247	0.1233	0.0216	0.0032	0.0427	-0.0017
$direct_1$	0.0015	-0.0087	0.0029	0.0032	-0.0022	-0.0054	0.0000	-0.0006	0.0037
$direct_2$	0.0060	-0.0028	-0.0104	0.0109	0.0051	0.0022	0.0011	-0.0000	-0.0004
$direct_3$	-0.0026	-0.0130	-0.0017	-0.0077	0.0020	0.0038	0.0014	0.0010	0.0008
$indirect_1$	0.0007	0.0394	0.3380	-0.0047	0.0277	0.1347	0.0010	0.0121	0.0416
$indirect_2$	-0.0022	0.0013	-0.0207	-0.0045	0.0016	0.0029	-0.0004	0.0002	0.0000
$indirect_3$	0.0001	-0.0449	-0.3340	0.0067	-0.0261	-0.1367	-0.0016	-0.0120	-0.0355
$total_1$	0.0022	0.0307	0.3409	-0.0015	0.0255	0.1294	0.0011	0.0115	0.0453
$total_2$	0.0038	-0.0015	-0.0311	0.0064	0.0066	0.0051	0.0007	0.0001	-0.0003
$total_3$	-0.0025	-0.0579	-0.3357	-0.0011	-0.0241	-0.1330	-0.0002	-0.0110	-0.0347
σ^2	0.0331	0.0094	-0.0116	-0.0106	-0.0006	0.0362	-0.0007	0.0071	-0.0006

Table 5: Mean-squared error results for 500 trials

	$N = 500$			$N = 1,000$			$N = 5,000$		
$\sigma^2 = 1$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$
Parameters	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE
ρ	0.0041	0.0040	0.0029	0.0025	0.0019	0.0016	0.0006	0.0004	0.0004
β_1	0.0018	0.0019	0.0018	0.0011	0.0010	0.0011	0.0002	0.0002	0.0002
β_2	0.0021	0.0021	0.0020	0.0009	0.0010	0.0010	0.0002	0.0002	0.0002
β_3	0.0022	0.0021	0.0021	0.0011	0.0011	0.0011	0.0002	0.0002	0.0002
γ_1	0.0032	0.0151	0.0022	0.0017	0.0104	0.0013	0.0004	0.0028	0.0003
γ_2	0.0037	0.0147	0.0024	0.0017	0.0129	0.0013	0.0004	0.0035	0.0003
γ_3	0.0029	0.0254	0.0022	0.0019	0.0124	0.0015	0.0006	0.0029	0.0005
$direct_1$	0.0020	0.0019	0.0021	0.0012	0.0010	0.0012	0.0002	0.0002	0.0003
$direct_2$	0.0023	0.0021	0.0023	0.0010	0.0010	0.0011	0.0002	0.0002	0.0002
$direct_3$	0.0024	0.0022	0.0024	0.0012	0.0011	0.0011	0.0002	0.0002	0.0002
$indirect_1$	0.0012	0.0116	0.1906	0.0007	0.0050	0.0749	0.0002	0.0010	0.0168
$indirect_2$	0.0004	0.0002	0.0058	0.0002	0.0001	0.0024	0.0000	0.0000	0.0004
$indirect_3$	0.0014	0.0113	0.1957	0.0007	0.0052	0.0717	0.0002	0.0010	0.0176
$total_1$	0.0015	0.0156	0.2025	0.0009	0.0064	0.0822	0.0002	0.0013	0.0177
$total_2$	0.0008	0.0033	0.0152	0.0003	0.0016	0.0067	0.0001	0.0003	0.0012
$total_3$	0.0015	0.0145	0.2111	0.0009	0.0073	0.0762	0.0002	0.0014	0.0190
σ^2	0.0041	0.0038	0.0041	0.0021	0.0018	0.0018	0.0005	0.0004	0.0004
	$N = 500$			$N = 1,000$			$N = 5,000$		
$\sigma^2 = 10$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$	$\rho = -0.6$	$\rho = 0.2$	$\rho = 0.6$
Parameters	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE
ρ	0.0081	0.0080	0.0052	0.0040	0.0040	0.0030	0.0011	0.0009	0.0008
β_1	0.0187	0.0165	0.0190	0.0103	0.0086	0.0103	0.0021	0.0022	0.0023
β_2	0.0196	0.0177	0.0196	0.0098	0.0095	0.0088	0.0019	0.0022	0.0021
β_3	0.0171	0.0197	0.0198	0.0100	0.0108	0.0102	0.0019	0.0020	0.0021
γ_1	0.0062	0.0222	0.0039	0.0028	0.0155	0.0021	0.0008	0.0048	0.0006
γ_2	0.0080	0.0128	0.0047	0.0035	0.0140	0.0026	0.0008	0.0055	0.0005
γ_3	0.0068	0.0370	0.0041	0.0031	0.0209	0.0025	0.0010	0.0048	0.0008
$direct_1$	0.0202	0.0168	0.0214	0.0111	0.0088	0.0113	0.0023	0.0022	0.0025
$direct_2$	0.0213	0.0180	0.0218	0.0107	0.0096	0.0099	0.0021	0.0022	0.0023
$direct_3$	0.0188	0.0200	0.0220	0.0109	0.0109	0.0113	0.0021	0.0020	0.0023
$indirect_1$	0.0049	0.0269	0.5831	0.0026	0.0118	0.1845	0.0006	0.0023	0.0386
$indirect_2$	0.0034	0.0022	0.0636	0.0018	0.0008	0.0257	0.0003	0.0002	0.0046
$indirect_3$	0.0051	0.0279	0.6054	0.0027	0.0112	0.1882	0.0006	0.0023	0.0383
$total_1$	0.0089	0.0550	0.6805	0.0048	0.0271	0.2247	0.0011	0.0057	0.0485
$total_2$	0.0077	0.0312	0.1558	0.0038	0.0157	0.0662	0.0007	0.0035	0.0134
$total_3$	0.0078	0.0614	0.7025	0.0045	0.0267	0.2284	0.0009	0.0054	0.0476
σ^2	0.4170	0.3584	0.4157	0.2220	0.2176	0.2015	0.0394	0.0411	0.0432

4 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).[9] The initial database includes 112,830 houses (with identified coordinates and having removed missing data) but we only consider homes with at east 1 and less than 9 bedrooms and baths, more than 200 square foot of living area and more than 500 square foot lotsize leading to a sample size of 110,984. Further, the sample has been checked to ensure the presence of twenty nearest neighboring homes within two miles of each home that had the same number of bedrooms, same number of full plus half baths and same age category, leading to the final sample size.[10]

The dependent variable is the *(logged) selling price*, with two explanatory variables: *log(total living area)* (in square feet), *log(lot size)* (in square feet). 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 as a way to approximate prices of comparable homes, and then treats house characteristics such as bedrooms, baths and house age as resulting in an adjustment to selling prices. For example, an additional bath room or bedroom adds some dollar amount to the selling price, while a home that is one year older would adjust selling price downward. This type of model specification envisions that these individual house characteristics can be adjusted in a ceteris paribus fashion, each contributing to a partial derivative impact on selling price.

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 buyers' identify comparable homes in their search for a home. 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, if our approach is successful, we would expect to see a higher level of dependence

[9] The dataset is publicly avaliable at: `http://homepages.uc.edu/~brasindd/housing.html`

[10] 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.

between homes and the autoregressive lag of selling prices based on the convex combination. Intuitively, a convex combination based on houses with similar characteristics should better capture selling prices of truly comparable nearby homes. This is because comparability reflects design aspects of homes that buyers care about, not merely selling prices based purely on proximate location. 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. This is because homes that are comparable in design reflect the true basis for buyers relative assessment of willingness to pay comparable prices for comparable 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.

Three connectivity matrices were constructed to reflect the nearest neighboring homes that potential buyers would view as similar along three dimensions. Specifically, the nearest homes (within two miles) with the same number of bedrooms (W_{beds}), the nearest neighboring homes with the same number of full plus half-baths (W_{baths}), and the nearest homes in the same age category (W_{age}). The notion is that buyers interested in a home (say i) would consider a nearby home (say j) having the same number of bedrooms, same number of baths or similar age/vintage as *comparable* in design. Buyers would view homes nearby that did not have a similar number of beds, baths or house age in their search for a home as not truly comparable.

Of course, there is the question of how many nearest neighboring homes should be used to construct the convex combination of connectivity matrices $W_c = \gamma_1 W_{beds} + \gamma_2 W_{baths} + (1 - \gamma_1 - \gamma_2) W_{age}$, which can be answered using estimated log-marginal likelihoods and associated model probabilities. Table 6 shows the results of these calculations, which point to a connectivity matrix W_c based on 13 and 14 nearest neighbors, with posterior model probabilities around 0.034 and 0.966 for each of these, respectively. Hence, in this application, the model uncertainty is quite low.

A question of interest is how similar connectivity matrices based on alternative numbers of nearest neighbors such as 13 and 14 nearest homes having the same number of bedrooms, baths and age category are. 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 (ν) to produce vectors $W_{beds}\nu$, $W_{baths}\nu$, $W_{age}\nu$. The correlation between these vectors can then be used to judge similarity.

Table 6: Model probabilities for models specified using W_c matrices based on 1 to 20 nearest neighboring homes

# neighbors	log-marginal likelihood	Model Probabilities
1	-34948.997	0.0000
2	-30087.991	0.0000
3	-27616.608	0.0000
4	-26277.835	0.0000
5	-25382.597	0.0000
6	-24863.930	0.0000
7	-24520.929	0.0000
8	-24273.111	0.0000
9	-24105.784	0.0000
10	-23988.988	0.0000
11	-23934.369	0.0000
12	-23914.473	0.0000
13	-23902.279	0.0335
14	-23898.918	0.9665
15	-23936.593	0.0000
16	-23980.722	0.0000
17	-24041.366	0.0000
18	-24088.802	0.0000
19	-24143.812	0.0000
20	-24194.471	0.0000

Two correlation matrices based on 13 and 14 nearest neighboring homes are shown in Table 7, where we see correlations around 0.5 for spatial lag vectors $W_{beds}\nu$ and $W_{baths}\nu$ and a slightly higher correlation of 0.55 between spatial lag vectors $W_{baths}\nu$ and $W_{age}\nu$. Hence,the neighbors based on common number of bedrooms, baths or of similar age are not the same.

A related question is the similarity of these types of connectivity matrices with the conventional spatial connectivity matrix based on 13 or 14 geographically nearest neighboring homes. Correlations between the spatial lag vector $W_{space}\nu$ and those constructed using 13 and 14 nearest neighboring homes with the same number of bedrooms, baths and age category are shown in Table 8. 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

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

13 neighbors	W_{beds}	W_{baths}	W_{age}
W_{beds}	1.0000		
W_{baths}	0.5067	1.0000	
W_{age}	0.5264	0.5529	1.0000
14 neighbors	W_{beds}	W_{baths}	W_{age}
W_{beds}	1.0000		
W_{baths}	0.5078	1.0000	
W_{age}	0.5256	0.5510	1.0000

the same vintage/age, which is not surprising given that homes in a given neighborhood are frequently built around the same time. These results point to the inherent truth that homes located nearby tend to be similar in design, which of course accounts for the past success of spatial lags of house prices working well in spatial autoregressive hedonic house price regressions to approximate comparable homes.

Table 8: Correlation of $W_{space}\nu$, $W_{beds}\nu$, $W_{baths}\nu$ $W_{age}\nu$

13 neighbors	W_{space}	W_{beds}	W_{baths}	W_{age}
W_{space}	1.0000			
W_{beds}	0.5880	1.0000		
W_{baths}	0.6425	0.5067	1.0000	
W_{age}	0.7420	0.5264	0.5529	1.0000
14 neighbors	W_{space}	W_{beds}	W_{baths}	W_{age}
W_{space}	1.0000			
W_{beds}	0.5873	1.0000		
W_{baths}	0.6399	0.5078	1.0000	
W_{age}	0.7379	0.5256	0.5510	1.0000

Table 9 reports estimation results for a model based on the conventional spatial connectivity matrix constructed using 14 nearest homes along with estimates based on the convex combination of $W_c = \gamma_1 W_{beds} + \gamma_2 W_{baths} + (1 - \gamma_1 - \gamma_2)W_{age}$. Of course, this ignores posterior probability support of 3.35 percent for the model specification based on 13 nearest neighboring homes, about which we have more to say later.

Table 9: SAR model hedonic house price regressions (14 neighbors)

	W_{space}			$W_c = [W_{beds}, W_{baths}, W_{age}]$		
	Lower	Posterior	Upper	Lower	Posterior	Upper
Estimates (variable)	0.01	median	0.99	0.01	median	0.99
Constant	-0.4240	-0.3574	-0.3006	-0.5963	-0.5368	-0.4833
β(log(TLA))	0.3825	0.3910	0.3999	0.2973	0.3060	0.3146
β(log(lotsize))	0.0540	0.0591	0.0631	0.0587	0.0624	0.0663
ρ(Wy)	0.7322	0.7392	0.7456	0.7990	0.8057	0.8122
$\gamma_1(W_{beds})$		-		0.1708	0.1903	0.2077
$\gamma_2(W_{baths})$		-		0.3556	0.3762	0.3954
$\gamma_3(W_{age})$		-		0.4123	0.4339	0.4547
Direct(log(TLA))	0.4122	0.4216	0.4309	0.3196	0.3287	0.3376
Direct(log(lotsize))	0.0583	0.0637	0.0680	0.0631	0.0670	0.0712
Indirect(log(TLA))	1.0378	1.0768	1.1168	1.1837	1.2458	1.3050
Indirect(log(lotsize))	0.1505	0.1629	0.1741	0.2363	0.2539	0.2713
Total(log(TLA))	1.4539	1.4986	1.5462	1.5065	1.5747	1.6362
Total(log(lotsize))	0.2106	0.2267	0.2417	0.2991	0.3210	0.3415
Log-marginal likelihood		-26490.7439			-23898.918	

The results based on a conventional spatial weight matrix constructed using the nearest 14 homes and the convex combination model based on a convex combination of weights constructed using nearby homes having the same number of bedrooms, bathrooms and house age are different 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 3.29 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 $\rho = 0.8057$ for the convex combination model versus $\rho = 0.7392$ for the spatial neighbors model. Of course, this leads to larger indirect or spillover effects, associated with the impact of home prices from the set of more comparable homes identified by the convex combination model. As noted earlier, larger spillover estimates may also arise because the convex combination model weight matrix is denser than the spatial weight matrix.[11] It is also the case that the log-marginal likelihood shows an improvement for the convex combination model relative to that based on spatial proximity weights.

In the case of the elasticity response of selling price to changes in lotsize, we see no significant difference in the direct effect, but there is a significant difference in the indirect effects. Here again, given the higher level of dependence, we would expect to see a larger indirect effect of lotsizes from more comparable homes on the selling prices. The significant difference in indirect lotsize effects estimates from the two specifications lead to a significant difference in the total effect of lotsize for the two specifications, with those from the convex combination model being larger.

Accounting for model uncertainty

The estimates presented in the previous section rely on connectivity matrix specifications constructed from 14 nearest neighboring homes, whereas the log-marginal likelihood calculations showed support for models based on both 13 and 14 nearest neighbors. The Bayesian solution to this type of model uncertainty is to produced estimates based on both model specifications and weight the estimates using posterior model probabilities. Using a convex combination of six connectivity matrices, three based on W_{beds}, W_{baths} and W_{age} formed using 13 nearest neighboring homes, and another three based on 14 nearest neighboring homes comparable in these characteristics, we have $M = 2^L - L - 1 = 57$ different models.

Table 10 shows log-marginal likelihoods for 14 of these 57 models that have posterior probabilities greater than 0.001. Estimates for the parameters γ_1 to γ_6 associated with each of these models are also shown in the table,

[11] A check of non-zero off-diagonal elements from the two weight matrices showed this was indeed the case. For example, in the case of a 14 neighbors convex combination weight matrix the average number of non-zero off-diagonal elements was 25, versus 14 for the spatial weight matrix based on 14 nearest neighbors.

along with model averaged estimates for these parameters.

Table 10: Model probabilities for models specified using W_c matrices based on 13 and 14 nearest neighboring homes

Models	log-marginal likelihood	Model probs.	13 nearest neighbors			14 nearest neighbors		
			W_{beds}	W_{baths}	W_{age}	W_{beds}	W_{baths}	W_{age}
1	-23894.791	0.001	0.117	0.375	0.139	0.078	0.000	0.291
2	-23894.227	0.002	0.143	0.211	0.148	0.051	0.165	0.281
3	-23893.815	0.003	0.000	0.241	0.000	0.190	0.138	0.431
4	-23893.725	0.003	0.000	0.213	0.162	0.190	0.165	0.270
5	-23893.672	0.004	0.140	0.230	0.000	0.053	0.148	0.429
6	-23893.345	0.005	0.000	0.376	0.153	0.193	0.000	0.279
7	-23893.024	0.007	0.128	0.375	0.000	0.067	0.000	0.429
8	-23892.724	0.009	0.193	0.000	0.000	0.000	0.375	0.432
9	-23892.5	0.012	0.000	0.376	0.000	0.193	0.000	0.431
10	-23892.05	0.019	0.193	0.000	0.181	0.000	0.374	0.251
11	-23890.749	0.068	0.195	0.375	0.133	0.000	0.000	0.297
12	-23890.143	0.125	0.194	0.194	0.153	0.000	0.182	0.277
13	-23889.276	0.297	0.194	0.217	0.000	0.000	0.161	0.429
14	-23888.873	0.444	0.194	0.376	0.000	0.000	0.000	0.430
BMA	23889.542	-	0.188	0.293	0.033	0.005	0.083	0.399

From the table we see that the highest probability model assigns non-zero γ estimates to W_{age} based on 14 nearest neighboring homes, W_{baths} and W_{beds} based on 13 nearest neighboring homes, which accounts for 44.4% of the posterior probability mass. Model averaged estimates assign small probabilities to W_{age} based on 13 neighbors, and W_{beds} and W_{baths} based on 14 neighbors, accounting for model uncertainty regarding weights based on 13 and 14 nearest neighboring homes. The pattern of γ estimates is such that the largest weight is assigned to $W_{age}(\gamma = 0.399)$ based on 14 neighbors, then $W_{baths}(\gamma = 0.293)$ based on 13 neighbors and finally $W_{beds}(\gamma = 0.188)$ constructed from 13 neighbors. These results are in general agreement with those shown in Table 9, where the most weight is assigned to $W_{age}, (\gamma = 0.4339)$, then $W_{baths}(\gamma = 0.3762)$ and finally $W_{beds}(\gamma = 0.1903)$.

Table 11 shows a comparison of the convex combination SAR model estimates based on 14 neighboring homes and the BMA estimates that include both 13 and 14 neighboring homes. We see that estimates and inferences regarding the direct and indirect effects are very similar. further refinement of the nature of the weight matrix do not have a material impact on ultimate estimates and inferences, because conclusions drawn based on the model using only 14 nearest neighboring homes are robust to any model un-

Table 11: SAR convex model versus BMA convex model estimates

	14 neighbors			13 and 14 neighbors		
	Lower	Posterior	Upper	Lower	Posterior	Upper
Estimates(variable)	0.01	median	0.99	0.01	median	0.99
Constant	-0.5963	-0.5368	-0.4833	-0.5573	-0.5235	-0.4896
β(log(TLA))	0.2973	0.3060	0.3146	0.3013	0.3063	0.3110
β(log(lotsize))	0.0587	0.0624	0.0663	0.0602	0.0622	0.0645
ρ(Wy)	0.7990	0.8057	0.8122	0.8002	0.8044	0.8082
$\gamma_1(W_{beds,13NN})$		-		0.1772	0.1886	0.1992
$\gamma_2(W_{baths,13NN})$		-		0.2465	0.2934	0.3385
$\gamma_3(W_{age,13NN})$		-		0.0143	0.0336	0.0517
$\gamma_4(W_{beds,14NN})$	0.1708	0.1903	0.2077	0.0048	0.0054	0.0063
$\gamma_5(W_{baths,14NN})$	0.3556	0.3762	0.3954	0.0377	0.0834	0.1290
$\gamma_6(W_{age,14NN})$	0.4123	0.4339	0.4547	0.3747	0.3965	0.4186
Direct(log(TLA))	0.3196	0.3287	0.3376	0.3245	0.3297	0.3348
Direct(log(lotsize))	0.0631	0.0670	0.0712	0.0648	0.0669	0.0694
Indirect(log(TLA))	1.1837	1.2458	1.3050	1.2062	1.2361	1.2654
Indirect(log(lotsize))	0.2363	0.2539	0.2713	0.2421	0.2510	0.2612
Total(log(TLA))	1.5065	1.5747	1.6362	1.5344	1.5658	1.5992
Total(log(lotsize))	0.2991	0.3210	0.3415	0.3074	0.3179	0.3307
Log-marginal likelihood		-23898.918			-23889.1542	

certainty regarding use of 13 or 14 nearest neighbors. This type of finding is consistent with what LeSage and Pace (2014) label *The biggest myth in spatial econometrics*, estimates and inferences should not change dramatically when (small) refinements of the weight structure are made to a weight matrix that exhibits good model fit. This result is strenghtened in our case as model uncertainty is very small.

5 Conclusion

We consider estimating spatial regression models that utilize convex combinations of connectivity structures that expand on conventional approaches based only on spatial proximity of observations. Debarsy and LeSage (2018) argue that models constructed using a convex combination of weight matrices to form a single linear combination of alternative weight structures hold intuitive appeal. These models allow us to extend the types of cross-sectional dependence modeled using spatial regression methods, beyond conventional spatial dependence.

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.

We focus on challenges for estimation and inference that arise for this type of model where the matrix W_c is a function of estimated parameters γ_ℓ. One is that the log-determinant term that arises in the likelihood cannot be pre-calculated over a range of values for the spatial dependence parameter as is conventionally done in single weight matrix spatial regression models. A second issue relates to dealing with the restrictions imposed on the parameters γ_ℓ: $0 \leq \gamma_\ell \leq 1, \ell = 1, \ldots, L$ and $\sum_{\ell=1}^{L} \gamma_\ell = 1$. A final challenge arises when calculating measures of dispersion for the partial derivatives $\partial y/\partial x$ that LeSage and Pace (2009) label *effects estimates*. An empirical measure of dispersion for the effects is typically constructed by evaluating the partial derivatives using a large number (say 1,000) of draws for the parameters. The matrix expression for the partial derivatives involves the inverse of an $n \times n$ matrix for which LeSage and Pace (2009) propose a trace approximation that estimates main diagonal elements of the matrix inverse required to calculate effects estimates.

We modify earlier work by Pace and LeSage (2002) that relies on a Taylor series approximation to the log-determinant involving matrices of pre-calculated traces based on the L underlying weight matrices to address the first issue. The second issue is resolved using a reversible jump MCMC block sampling scheme for the parameters $\gamma_\ell, \ell = 1, \ldots, L$ that proposes a block of candidate values for γ_ℓ that obey the restrictions. The block of parameters is accepted or rejected in a Metropolis-Hastings step, meaning that the parameters γ_ℓ always obey the restrictions. The third problem is solved using exact pre-calculated traces in conjunction with stochastic estimates of higher-order traces, which allows for an extension of the method from LeSage and Pace (2009) that avoids calculation of the matrix inverse.

Our computational approach allows a large number of MCMC draws to be carried out in very little time. Computational speed is required when estimating these models because a large number of draws are required to adequately estimate the parameters that act as weights in the linear combination of matrices, due to low Metropolis-Hastings acceptance rates.

Results from a Monte Carlo study show that our method produces estimates with small bias and mean-squared errors, as well as good coverage for the parameters as well as the direct and indirect effects. LeSage and Pace (2017) point out that past Monte Carlo studies in spatial econometrics have largely ignored coverage of true values of the parameters by the empirical measures of dispersion used for inference regarding the direct and indirect effects.

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 compare house prices.

References

Barry, R.P., R.K. Pace (1999), A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices, *Linear Algebra and its Applications*, 289, 41–54.

Blankmeyer, E., J.P. LeSage, J.R. Stutzman, K.J. Knox, R.K. Pace (2011), Peer group dependence in salary benchmarking: A statistical model, *Managerial and Decision Economics*, 32(2), 91–104.

Brasington,D.M., 2007, Private Schools and the Willingness to Pay for Public Schooling, *Education Finance and Policy*, 2, 152–174. February.

Brasington D.M., D.R. Haurin, 2006, Educational Outcomes and House Values: A Test of the value added Approach, *Journal of Regional Science*, 46, 245–268.

Brasington, D.M., D. Hite, 2008, A mixed index approach to identifying hedonic price models, *Regional Science and Urban Economics*, 38, 271–284.

Debarsy N., J.P. LeSage (2018), Flexible dependence modeling using convex combinations of different types of connectivity structures, *Regional Science and Urban Economics* , 69, 48–68.

Dennison D.G.T., C.C. Holmes, B.K. Mallick , A.F.M. Smith (2002), *Bayesian methods for nonlinear classification and regression*, West Sussex, England, John Wiley & Sons.

Dittrich, D., R. Th.A.J. Leenders, J. Mulder (2017), Bayesian estimation of the network autocorrelation model, *Social Networks*, 48:1, 213–236.

Green, P. J. (1995), Reversible jump markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, 711–732.

Golub, G.H., C.F. Van Loan (1996), *Matrix Computation, Third Edition*, The Johns Hopkins University Press, Baltimore.

Hazir, C.S., J.P. LeSage, C. Autant-Bernard (2018), The role of R&D collaboration networks on regional innovation performance, *Papers in Regional Science*, 97, 549–567.

Hepple, L.W. (1995a), Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds, *Environment and Planning A*, 27:3, 447–469.

Hepple, L.W. (1995b), Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms, *Environment and Planning A*, 27:4, 615–644.

LeSage, J. P., R.K. Pace (2009), *Introduction to Spatial Econometrics*, (Boca Raton: Taylor Francis/CRC Press).

LeSage, J.P., R.K. Pace (2011), Pitfalls in higher order model extensions of basic spatial regression methodology, *Review of Regional Studies*, 41(1), 13–26.

LeSage, J.P., R.K. Pace (2014), The biggest myth in spatial econometrics, *Econometrics* 2, 217–249.

LeSage, J.P., R.K. Pace (2017), Spatial econometric Monte Carlo studies: raising the bar, *Empirical Economics* https://doi.org/10.1007/s00181-017-1330-6

Pace, R.K., R.P. Barry, O.W. Gilley, C.F. Sirmans, 2000, A method for spatial-temporal forecasting with an application to real estate prices, *International Journal of Forecasting*, 16, 229–246.

Pace, R.K., J.P. LeSage (2002), Semiparametric maximum likelihood estimates of spatial dependence, *Geographical Analysis*, 34(1), 75–90.

Pace, R.K., C.F. Sirmans, V.C. Slawson, 2002, Are Appraisers Statisticians?, in *Real Estate Valuation Theory*, Wang, K. and Wolverton, M.L. (Eds.), Springer, 31–45.

Richardson, S., P.J. Green (1997), On bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society Series B*, 59, 731–792.

Zellner, A. (1971), *Introduction to Bayesian Inference in Econometrics* (John Wiley, New York).

Figures

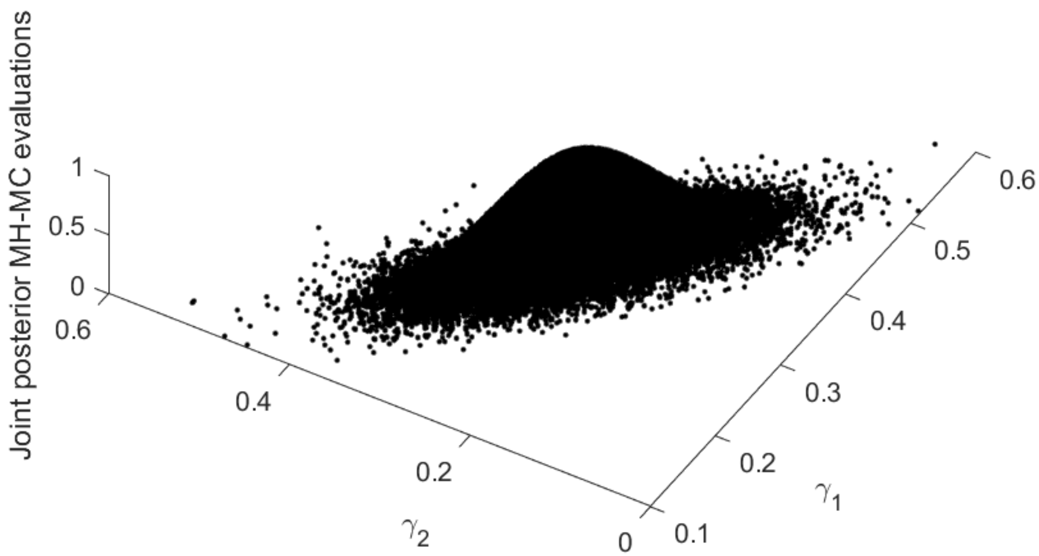

Figure 1: Joint posterior points for γ_1, γ_2

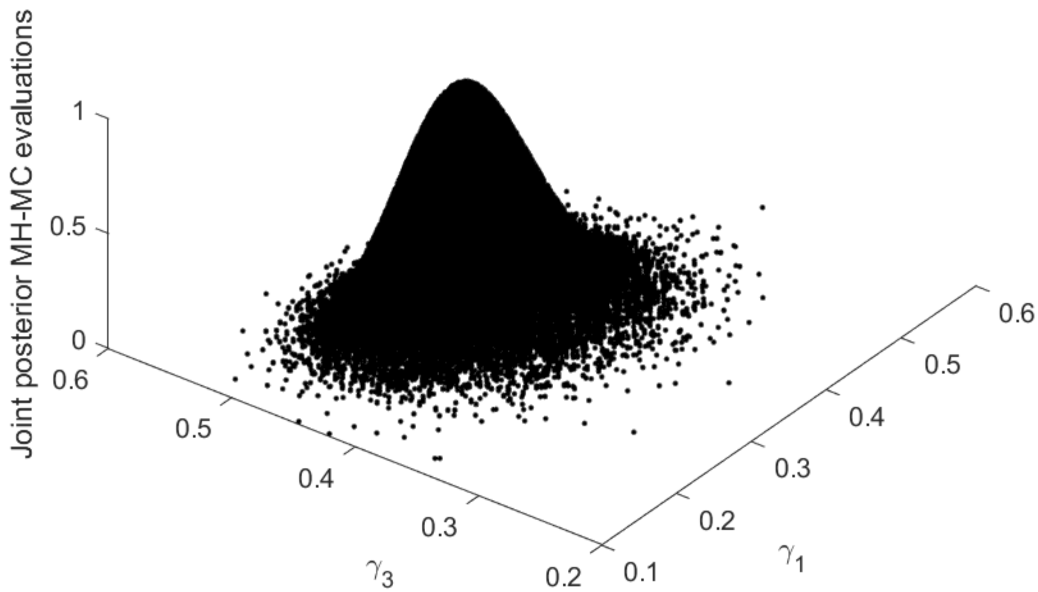

Figure 2: Joint posterior points for γ_1, γ_3

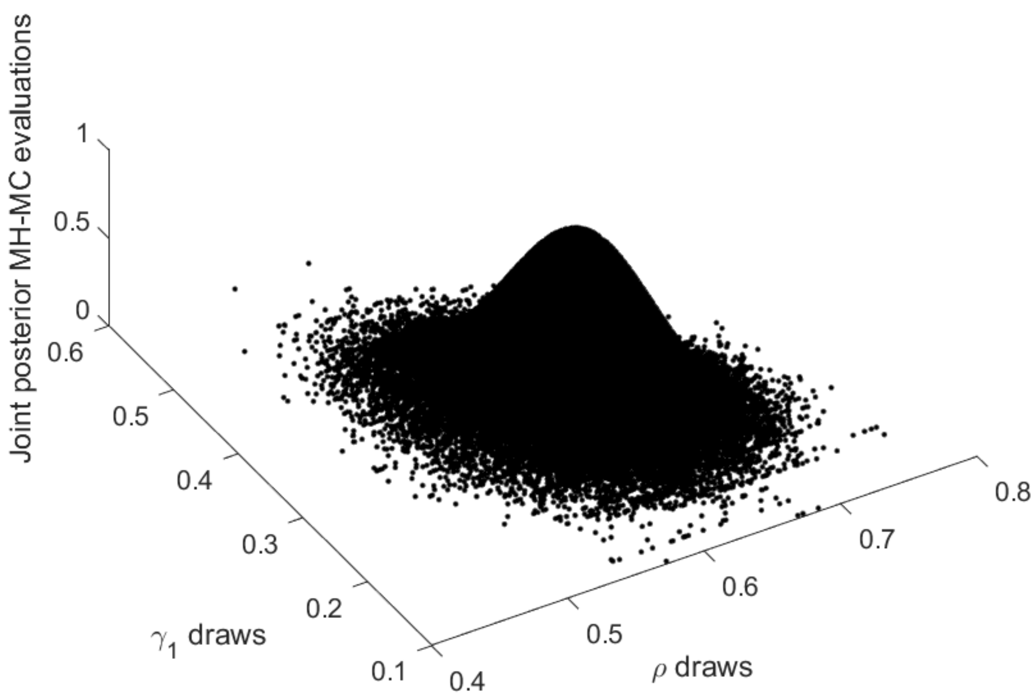

Figure 3: Joint posterior points for γ_1, ρ

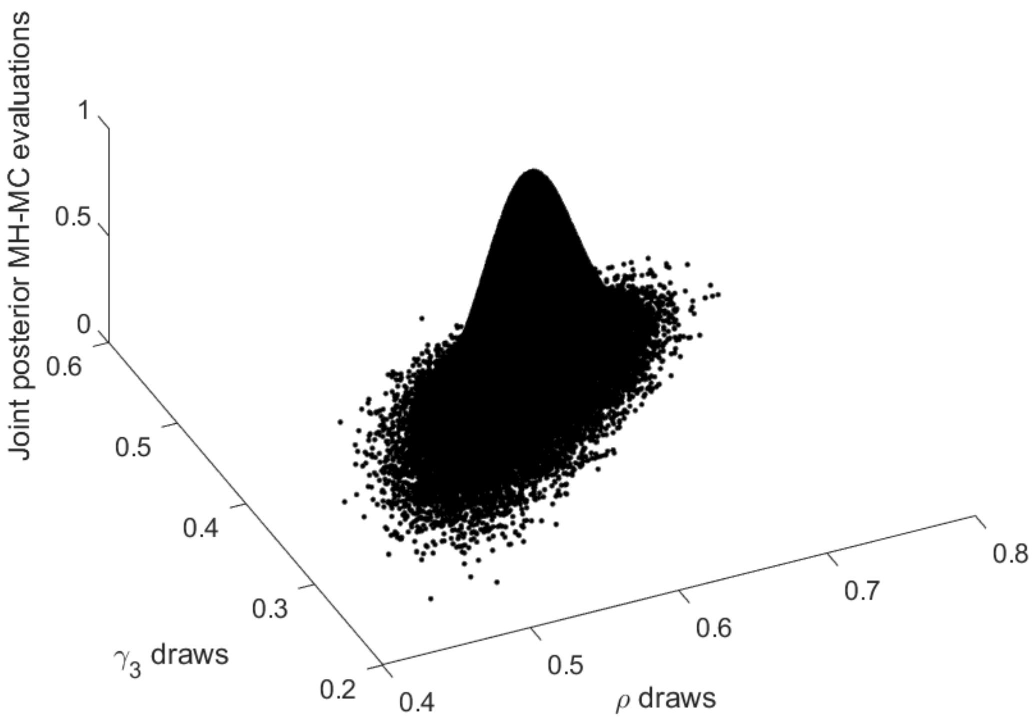

Figure 4: Joint posterior points for γ_3, ρ