Bayesian Model Averaging for Spatial Econometric Models

James P. LeSage,[1] Olivier Parent[2]

[1]McCoy Endowed Chair of Urban and Regional Economics, McCoy College of Business Administration, Department of Finance and Economics, Texas State University—San Marcos, San Marcos, TX, [2]Department of Economics, University of Cincinnati, Cincinnati, OH

We extend the literature on Bayesian model comparison for ordinary least-squares regression models to include spatial autoregressive and spatial error models. Our focus is on comparing models that consist of different matrices of explanatory variables. A Markov Chain Monte Carlo model composition methodology labeled MC³ by Madigan and York is developed for two types of spatial econometric models that are frequently used in the literature. The methodology deals with cases where the number of possible models based on different combinations of candidate explanatory variables is large enough such that calculation of posterior probabilities for all models is difficult or infeasible. Estimates and inferences are produced by averaging over models using the posterior model probabilities as weights, a procedure known as Bayesian model averaging. We illustrate the methods using a spatial econometric model of origin–destination population migration flows between the 48 U.S. states and the District of Columbia during the 1990–2000 period.

Introduction

There is a great deal of literature on Bayesian model comparison for nonspatial regression models, where alternative models consist of those based on differing matrices of explanatory variables. For example, Zellner (1971) sets forth the basic Bayesian theory behind model comparison for the case where a discrete set of m alternative models are under consideration. The approach involves specifying prior probabilities for each model as well as prior distributions for the regression parameters. Posterior model probabilities are then calculated and used for inferences regarding the alternative models based on different sets of explanatory variables.

Correspondence: James P. LeSage, McCoy Endowed Chair of Urban and Regional Economics, McCoy College of Business Administration, Department of Finance and Economics, Texas State University—San Marcos, San Marcos, TX 78666
e-mail: jlesage@spatial-econometrics.com

More recent works such as that by Fernández, Ley, and Steel (2001a, b) consider cases where the number of possible models m is large enough such that calculation of posterior probabilities for all models is difficult or infeasible. A Markov Chain Monte Carlo model composition methodology known as MC^3 proposed by Madigan and York (1995) has gained popularity in the mathematical statistics and econometrics literature (e.g., Denison, Mallick, and Smith 1998; Raftery, Madigan, and Hoeting 1997; Fernández, Ley, and Steel 2001a, b). The popularity of MC^3 arises in part from its ability to provide a theoretically justifiable approach to a question that often arises in regression modeling—which explanatory variables are most important in explaining variation in the dependent variable vector?

This article develops the MC^3 methodology for two important spatial econometric regression models that have received widespread application (LeSage and Pace 2004). A host of additional complications arise when attempting to extend regression-based approaches to these spatial regression estimators. We provide theoretical details regarding these issues as well as computationally efficient solutions.

The models for which we provide an MC^3 modeling approach are members of a class of spatial regression models introduced in Ord (1975) and elaborated in Anselin (1988), shown in (1) and (2). The sample of n observations in the vector y represents a cross-section of regions located in space, for example, counties, states, or countries.

$$y = \rho W y + \alpha \iota_n + X\beta + \varepsilon \tag{1}$$

$$y = \alpha \iota_n + X\beta + u \tag{2}$$

$$u = \lambda W u + \varepsilon$$

The n by k matrix X contains explanatory variables as in ordinary least-squares (OLS) regression, β is an associated k by 1 parameter vector, and ε is an n by 1 disturbance vector, which we assume takes the form $\varepsilon \sim N(0, \sigma^2 I_n)$. The scalar α represents the intercept parameter and ι_n an n by 1 vector of ones. The n by n matrix W specifies the structure of spatial dependence between observations (y) or disturbances (u), with a common specification having elements $W_{ij} > 0$ for observations $j = 1, \ldots, n$ sufficiently close (as measured by some distance metric) to observation i. As noted above, observations reflect geographical regions, and so distances might be calculated on the basis of the centroid of the regions/observations. The expressions Wy and Wu produce vectors that are often called spatial lags, and ρ and λ denote scalar parameters to be estimated along with β and σ^2. Nonzero values for the scalars ρ and λ indicate that the spatial lags exert an influence on y or u. In this study, we refer to the model in (1) as the spatial autoregressive model (SAR) and that in (2) as the spatial error model (SEM). We note that setting $\rho = 0$ in the SAR model or $\lambda = 0$ in the SEM model produces a least-squares regression model.

There are a number of alternative ways to specify the spatial connectivity structure for the matrix W, where a particular specification selects elements i, j that are nonzero, reflecting spatial dependence between observations i and j. By convention, the diagonal elements of the spatial weight matrix W are set to zero to preclude an observation or disturbance term from dependence on its own value. We do not focus on specification issues related to the weight matrix W, a potentially important aspect of SAR and SEM model specification. We focus exclusively on model specification issues regarding the choice of explanatory variables that appear in the matrix X. An implication of this is that estimates for β, ρ, λ, and σ are conditional on the particular weight structure used, but this is implicitly assumed by conventional maximum likelihood estimation algorithms for the models in (1) and (2), which treat the weight matrix W as given or fixed.

A related model specification issue that arises in applied work is selection of the appropriate model, SAR or SEM in our case.[1] Numerous Monte Carlo studies of alternative systematic or sequential specification search approaches that address this facet of spatial regression model specification have been published in the literature, and a review of these can be found in Florax, Folmer, and Rey (2003).

The literature focusing on the specific question of which explanatory variables should be included in the X matrix for the class of models including SAR and SEM models is less prevalent. Hepple (1995a, b) sets forth conventional Bayesian methods for the case involving a small set of m models, where it is feasible to calculate m posterior model probabilities. To our knowledge, the topic of MC^3 methodology for this class of models has not yet been discussed.

In the next section, we review a unified Bayesian approach to model specification in the context of the SAR and SEM models. We extend the discussion set forth in Hepple (1995a, b) by including informative priors for the parameters β, λ, ρ, and σ in these models. Our presentation in the third section focuses on the MC^3 methodology and details needed to apply this approach to the SAR and SEM models. The fourth section of the article provides an illustrative example of the proposed method in a model of origin–destination migration flows across the 48 contiguous U.S. states and the District of Columbia. There are numerous other spatial econometric applications where interest focuses on which explanatory variables are most important, providing a large number of opportunities to apply the methods described here.

A Bayesian approach to model specification

Other authors have set forth the Bayesian theory behind model comparison that involves specifying prior probabilities for each of the m alternative models $M = M_1$, $M_2, \ldots, M_m$ under consideration, which we label $\pi(M_i)$, $i = 1, \ldots, m$, as well as prior distributions for the parameters $\pi(\eta)$, where $\eta = (\rho, \alpha, \beta, \sigma)$ (e.g., Zellner 1971; Fernández, Ley, and Steel 2001b). Our focus here is on comparing models with different explanatory variables.

If the sample data are to determine the posterior model probabilities, the prior probabilities should be set to equal values of $1/m$, making each model equally likely a priori. These are combined with the likelihood for y conditional on η as well as the set of models M, which we denote $p(y|\eta, M)$. The joint probability for M, η, and y takes the form

$$p(M, \eta, y) = \pi(M)\pi(\eta|M)p(y|\eta, M) \tag{3}$$

Application of Bayes rule produces the joint posterior for both models and parameters as

$$p(M, \eta|y) = \frac{\pi(M)\pi(\eta|M)p(y|\eta, M)}{p(y)} \tag{4}$$

The posterior probabilities regarding the models take the form

$$p(M|y) = \int p(M, \eta|y)\mathrm{d}\eta. \tag{5}$$

which requires integration over the parameter vector η. Our focus in the next two subsections is development of the marginal posterior in (5) for the SAR and SEM models as this plays a key role in the MC^3 model comparison methodology set forth in the following section.

Log-marginal posterior for the SAR model

We initially consider the SAR model, where the likelihood function for the parameters $\eta = (\alpha, \beta, \sigma, \rho)$, based on the data y takes the form shown in (6), where we include the spatial weight matrix W to indicate that the likelihood is conditional on the particular weight matrix used in the model. That is, the weight matrix is taken as given and treated in the same manner as the sample data information in y, X.

$$L(\eta; y, X, W) \propto (\sigma^2)^{-n/2}|I_n - \rho W|^{1/2} \exp\left\{-\frac{1}{2\sigma^2}e'e\right\} \tag{6}$$
$$e = (I_n - \rho W)y - \alpha\iota_n - X\beta$$

We can define the prior distributions for the parameters in η using a number of different approaches. An issue that arises in Bayesian model comparison is that posterior model probabilities can be sensitive to alternative specifications for the prior information. Use of diffuse priors on the model parameters might seem desirable in this situation, but can lead to paradoxical outcomes as noted by Lindley (1957). We draw on the work of Fernández, Ley, and Steel (2001b) for the case of least-squares models and rely on Zellner's g-prior for the parameters β in the model. An uninformative prior is assigned to the intercept parameter α and a conventional gamma prior is used for the parameter σ. We discuss two possible priors for the spatial dependence parameter ρ. In the sequel, we provide details regarding the priors for the parameters α, β, σ, and ρ in turn.

A great deal of computational simplicity arises if we follow Fernández, Ley, and Steel (2001a) and employ Zellner's g-prior (Zellner 1986) for the parameters β in the SAR model. In addition to simplifying matters, Fernández, Ley, and Steel (2001a) provide a theoretical justification for use of the g-prior as well as Monte Carlo evidence comparing nine alternative approaches to setting the hyperparameter g. Based on the Monte Carlo experiments, they recommend setting $g_i = 1/\max\{n, k_{M_i}^2\}$, where n denotes the number of observations, and k_{M_i} the number of explanatory variables in the least-squares matrix X associated with the model indexed by M_i.

We allow for an intercept parameter α in the model as the dependent variable vector y is not transformed. The matrix X represents the set of explanatory variables excluding the constant term, transformed by subtracting the means of each variable vector and dividing by the standard deviation. We assume that the spatial lag vector W_y appears in all models as does the intercept term, leaving only the variable vectors in the matrix X subject to change as we compare alternative models. This approach mirrors that developed by Fernández, Ley, and Steel (2001a), where the intercept term appears in all models. Another justification for this approach is that in the absence of spatial dependence where the parameter ρ equals zero, we would rely on standard regression-based models of the type described in Fernández, Ley, and Steel (2001a). In our spatial context, it seems intuitively appealing that, in the absence of explanatory variables X, the dependent variable y is modeled to follow a simple spatial autoregression, $y = \rho W y + \alpha 1_n + \varepsilon$.

We can adopt an approach similar to that of Fernández, Ley, and Steel (2001a), who rely on a noninformative prior for the intercept term α, which seems desirable as the vector y is untransformed. The g-prior on the regression coefficients for model M_i, which we designate as β_{M_i}, takes the form shown in (7). One motivation behind this prior setting is a desire to provide prior information that will not exert undue influence on posterior conclusions regarding choices between alternative models based on different sets of explanatory variables. Another aspect of the g-prior that is attractive in situations involving model comparisons based on alternative sets of explanatory variables is the ability of the before take the covariance structure of the explanatory variables in X into account.

$$\pi_b(\beta_{M_i}|\sigma^2) \sim N[\beta_0, \sigma^2(g_i X'_{M_i} X_{M_i})^{-1}] \tag{7}$$

Note that the studentized form of X will facilitate the prior mean of zero placed on β. Given the normal prior for the parameters β, an inverted gamma prior for σ^2 shown in (8) with parameters ν and $\bar{s}^2$ allows us to draw on forms from the conjugate nature of these two prior distributions from standard Bayesian regression theory.

$$\pi_s(\sigma^2) = \frac{(\nu\bar{s}^2/2)^{\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-(\frac{\nu+2}{2})}\exp(-\frac{\nu\bar{s}^2}{2\sigma^2}) \tag{8}$$

The results that we derive can be extended to the case of a noninformative prior on σ^2, as this arises as a special case when $\nu = \bar{s}^2 = 0$.

When considering a prior for the parameter ρ, we note that numerical integration will be required with respect to this parameter to produce the marginal posterior distribution for the SAR model. This allows flexibility in specifying a prior $\pi_r(\rho)$. However, we note that Lemma 2 in Sun, Tsutakawa, and Speckman (1999) indicates a restricted region of support for the parameter ρ. For row-standardized weight matrices W (where the row-sums are unity) typically used in application of these models, ρ must lie in the interval $[\mu_{min}^{-1}, 1]$ where μ_{min} denotes the minimum eigenvalue of the spatial weight matrix W (Lemma 2 in Sun, Tsutakawa, and Speckman 1999). Further, $\mu_{min} < 0$, so $\mu_{min}^{-1} < \rho < 1$. We also note that for many spatial data sets that exhibit significant positive spatial dependence, a restriction of $0 < \rho < 1$ may be appropriate. For more general cases where negative spatial dependence is not to be ruled out a priori, a restriction $-1 < \rho < 1$ is often reasonable in that this represents the effective region of support.

We suggest two alternative priors for the parameter ρ consistent with these notions: one being a uniform prior on the interval $[-1, 1]$ and the other a *Beta* (a_0, a_0) prior centered on zero.

$$
\begin{aligned}
\pi_r(\rho) &= U[-1, 1] \\
\pi_r(\rho) &= \frac{1}{Beta(a_0,\ a_0)} \frac{(1+\rho)^{a_0-1}(1-\rho)^{a_0-1}}{2^{2a_0-1}}
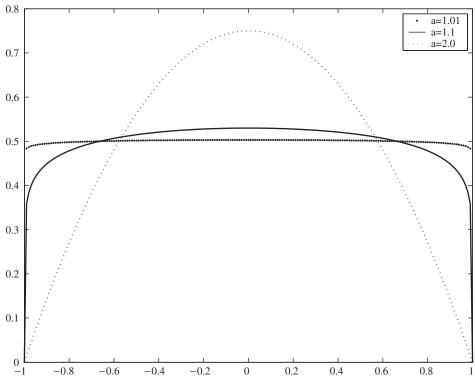\end{aligned}
\tag{9}
$$

Fig. 1 depicts prior distributions associated with hyperparameter values for the *Beta* prior of $a_0 = 1.01, 1.1$, and 2, illustrating that values of a_0 near unity produce a relatively uninformative prior that downweights to zero the prior weight placed on end points of the interval for ρ, consistent with theoretical restrictions. We note that this prior can also be used for the spatial dependence parameter λ in the SEM model discussed later.

Using Bayes' theorem, the marginal distribution of the SAR model can be written as

$$
\begin{aligned}
&\int \pi_b(\beta|\sigma^2)\pi_s(\sigma^2)\pi_r(\rho)p(y|\alpha, \beta, \rho, \sigma^2)d\beta d\sigma^2 d\rho \\
&= K_1(2\pi)^{-(n+k)/2}|C|^{1/2}\int |I_n - \rho W| \frac{1}{\sigma^{n+\nu+k+2}} \\
&\quad \times \exp\{-\frac{1}{2\sigma^2}[\nu\bar{s}^2 + S(\rho) + \beta'C\beta \\
&\quad + (\beta - \hat{\beta}(\rho))'(X'X)(\beta - \hat{\beta}(\rho))]\}\pi_r(\rho)d\beta d\sigma^2 d\rho,
\end{aligned}
\tag{10}
$$

Figure 1. Priors for ρ, λ.

with,

$$K_1 = \Gamma\left(\frac{v}{2}\right)^{-1}\left(\frac{v\bar{s}^2}{2}\right)^{v/2}$$

$$S(\rho) = e(\rho)'e(\rho)$$

$$e(\rho) = (I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n$$

$$C = gX'X$$

$$\hat{\beta}(\rho) = (X'X)^{-1}X'(I_n - \rho W)y$$

$$\hat{\alpha} = \bar{y} - \rho\overline{Wy}$$

$$\overline{Wy} = (1/n)\sum_i (Wy)_i$$

Using the properties of the multivariate normal probability density function (pdf) and the inverted gamma pdf to analytically integrate with respect to β and σ, we can arrive at an expression for the log marginal that will be required for model comparison purposes. We note that, as the intercept term is common to all models, this leads to $n-1$ as the degrees of freedom in the posterior

in the following equation

$$
p(\rho|y, X, W) = K_2 \left(\frac{g}{1+g}\right)^{k/2}
$$
$$
\times |I_n - \rho W|[v\bar{s}^2 + S(\rho) + Q(\rho)]^{-\frac{n+v-1}{2}} \pi_r(\rho)
$$

(11)

where,

$$
K_2 = \frac{\Gamma\left(\dfrac{n+v-1}{2}\right)}{\Gamma\left(\dfrac{v}{2}\right)}(v\bar{s}^2)^{\frac{v}{2}} \pi^{-\frac{n-1}{2}}
$$

$$
S(\rho) + Q(\rho) = \frac{1}{g+1}[(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n]'[(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n]
$$
$$
+ \frac{g}{g+1}[(I_n - \rho W)y - \hat{\alpha}\iota_n]'[(I_n - \rho W)y - \hat{\alpha}\iota_n]
$$

An important point regarding expression (11) is that we must rely on numerical integration to convert this into a scalar. This is in stark contrast to the results for conventional regression models, where analytical integration over the parameters β and σ leads to an expression that can be used to produce a scalar result. Details regarding computational issues surrounding numerical integration are set forth in Appendix A.

Log-marginal posterior for the SEM model

For the case of the SEM model, $y = \alpha\iota + X\beta + u$, $u = \lambda Wu + \varepsilon$, we also rely on a noninformative prior for the intercept term α and constrain this term to appear in all models. Again, the matrix X is transformed to studentized form, but not the vector y. As a prior mean of zero is assigned to the parameters β by the g-prior, it is important that the X matrix is well scaled. We rely on the normal-gamma priors for β, σ^2, but note that the matrix X does not contain a vector of ones, and the vector β does not contain an intercept term.

For this model, we define $y^* = y - \lambda Wy$ and $X^* = X - \lambda WX$, noting that X is studentized (or well scaled) to facilitate the prior mean of zero placed on the parameters β. A departure of the SEM from the SAR model is reliance on a g-prior based on the covariance matrix constructed from $X^{*\prime}X^*$, that is, $\sigma^2(gX^{*\prime}X^*)^{-1}$. The motivation for this is that the SEM estimates are: $\hat{\beta}(\lambda) = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$, and so following Zellner (1986), we should rely on the associated covariance matrix to form our g-prior. As in the case of the SAR model, this prior greatly simplifies the calculations needed to construct the log-marginal posterior for this model. It is convenient to define

$$
\bar{y}^* = \bar{y} - \lambda\overline{Wy}
$$
$$
\overline{Wy} = (1/n) \sum_i (Wy)_i
$$

which we use in (12) and the definitions that follow.

$$
p(\lambda|y, X, W) = K_2 \left(\frac{g}{1+g}\right)^{k/2}
$$
$$
\times |I_n - \lambda W|[v\bar{s}^2 + S(\lambda) + Q(\lambda)]^{-\frac{n+v-1}{2}} \pi_r(\lambda) \tag{12}
$$

where,

$$
S(\lambda) = \frac{1}{(1+g)} e^*(\lambda)' e^*(\lambda)
$$

$$
e^*(\lambda) = y^* - X^* \hat{\beta}(\lambda) - \hat{\alpha}(\lambda)\iota
$$

$$
Q(\lambda) = \frac{g}{(1+g)} (y^* - \bar{y}^*\iota)'(y^* - \bar{y}^*\iota)
$$

$$
\hat{\beta}(\lambda) = (X^{*'}X^*)^{-1} X^{*'} y^*
$$

$$
\hat{\alpha}(\lambda) = \bar{y}^* \tag{13}
$$

We note that when $\lambda = 0$, we have $y^* = y$ and $X^* = X$, and so our result is identical to that of Fernández, Ley, and Steel (2001a) in this case. As in the case of the SAR model, we must rely on numerical integration over the spatial dependence parameter λ to convert this into a scalar. Details regarding computational issues surrounding numerical integration are set forth in Appendix A.

MC^3 spatial model comparison

A large literature on Bayesian model averaging (BMA) over alternative linear regression models containing differing explanatory variables exists (Raftery, Madigan, and Hoeting 1997; Fernández, Ley, and Steel 2001a, b). The MC^3 approach introduced in Madigan and York (1995) is set forth here for the SAR and SEM models. For a regression model with k possible explanatory variables, there are 2^k possible ways to select regressors to be included or excluded from the model. For $k = 15$ say, we have 32,768 possible models, ruling out computation of the log-marginal for all possible models as impractical.

The MC^3 method of Madigan and York (1995) devises a strategic stochastic Markov chain process that can move through the potentially large model space and sample regions of high posterior support. This eliminates the need to consider all models by constructing a sampler that explores relevant parts of the very large model space. If we let M denote the current model state of the chain, models are proposed using a neighborhood, nbd(M), which consists of the model M itself along with models containing either one variable more (labeled a ''birth step'') or one variable less (a ''death step'') than M. A transition matrix, q, is defined by setting $q(M \rightarrow M') = 0$ for all $M' \notin$ nbd(M) and $q(M \rightarrow M')$ constant for all $M' \in$ nbd(M). The proposed model M' is compared with the current model state

M using the acceptance probability shown in following equation:

$$\min\left[1, \frac{p(M'|y)}{p(M|y)}\right] \qquad (14)$$

The use of univariate numerical integration methods described in Appendix A allows us to construct a Metropolis–Hastings sampling scheme that implements the MC^3 method. A vector of the log-marginal values for the current model M is stored during sampling along with a vector for the proposed model M'. These are then scaled and integrated to produce the ratio $p(M'|y)/p(M|y)$ in (14) that determines acceptance or rejection of the proposed model. In contrast to conventional regression models, there is a need to store log-marginal density vectors for each unique model found during the MCMC sampling to calculate posterior model probabilities over the set of all unique models visited by the sampler.

Although the use of birth and death processes in the context of Metropolis–Hastings sampling will theoretically produce samples from the correct posterior, Richardson and Green (1997), among others, advocate incorporating a "move step" in addition to the birth and death steps into the algorithm. We rely on this approach as there is evidence that combining these move steps improves the convergence of the sampling process (Denison, Mallick, and Smith 1998; Richardson and Green 1997). The move step takes the form of replacing a randomly chosen single variable in the current explanatory variables matrix with a randomly chosen variable not currently in the model. Specifically, we might propose a model with one less explanatory variable (death step) and then add an explanatory variable to this new model proposal (birth step). This leaves the resulting model proposal with the same dimension as the original one with a single component altered. This type of sampling process is often labeled "reversible jump" MCMC. The model proposals that result from birth, death, and move steps are all subjected to the Metropolis–Hastings accept/reject decision shown in (14), which is valid as long as the probabilities of birth, death, and move steps have an equal probability of 1/3.

The Bayesian solution to incorporating uncertainty regarding specification of the appropriate explanatory variables into the estimates and inferences is to "average" over alternative model specifications. This is in contrast with much applied work that relies on a single model specification identified using various model comparison criterion that lead to a "most preferred model." The averaging involves weighting alternative model specifications by their posterior model probabilities. We note that the MC^3 procedure identifies models associated with particular explanatory variables and assigns a posterior model probability to each of these models. Like all probabilities, the posterior model probabilities sum to unity, and so they can be used as weights to form a linear combination of estimates from models based on differing explanatory variables. This weighted combinations of sampling draws from the posterior

are used as the basis for posterior inference regarding the mean and dispersion of the individual parameter estimates.

Applied illustrations

In the first section, we contrast results from least-squares and SAR model MC^3 procedures using a data generated example based on a small data set and a small set of candidate explanatory variables. This allows us to validate our MC^3 approach in a setting where the true parameter values and model specification is known. In the second section, we apply the least-squares and SAR model MC^3 and BMA methods to a spatial sample of 2401 origin–destination migration flows from each of the 48 U.S. states plus the District of Columbia. As in most models of origin–destination flows, we have two sets of explanatory variables: one associated with characteristics of the origin regions/states and another with the destination characteristics. This provides an illustration of how the SAR model MC^3 and BMA methodology works in applied practice. As the sample data exhibits strong spatial dependence, we would expect differences between the least-squares and spatial results concerning which of the origin–destination characteristics are important in determining population migration flows.

A small sample illustration

If the data-generating process is the SAR model, then $\hat{\beta}_{SAR} = (X'X)^{-1}X'(I_n - \rho W)y$, and least-squares estimates for β are biased and inconsistent (see LeSage and Pace 2004). In these cases, we would expect that least-squares MC^3 procedures would not produce accurate estimates and inferences regarding which variables are important. If the data-generating process is the SEM model, then $\hat{\beta}_{SEM} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$, where $\tilde{X} = (X - \lambda WX), \tilde{y} = (y - \lambda Wy)$. Least-squares estimates are unbiased but inefficient (analogous to the case of serial correlation in the disturbances), and so we might expect better results for least-squares MC^3 procedures.

To illustrate differences between least-squares and spatial results, we used a 49-location data set from Anselin (1988) containing Columbus, Ohio neighborhoods, along with observations on the median housing values (*hvalue*) for each neighborhood and household income (*hincome*). These variable values were used as the basis for a data-generated experiment where the true parameter values as well as the true model specification will be known. The spatial weight matrix was based on the four nearest neighbors to each of the spatial observations, and the matrix was row-normalized to have row sums of unity. Two additional explanatory variables were constructed using spatial lags of the house values and household income. These variables represent an average of housing values and household income from the nearest four neighborhoods, constructed through multiplication of these variable vectors by the row-normalized spatial weight matrix W. Intuitively, housing values and household income levels in nearby neighborhoods might contribute to explaining variation in the *y* variable, which was neighborhood crime

rates in the Anselin (1988) application. This produces a model shown in the following equation:

$$y = \alpha\iota + \rho Wy + \beta_1 hvalue + \beta_2 hincome \\ + \beta_3 W \cdot hvalue + \beta_4 W \cdot hincome + \varepsilon \tag{15}$$

The intercept term and spatial lag are included in all models, and so the number of possible candidate variables is 4, leading to $2^4 = 16$ possible models. This makes it simple to validate our MC^3 algorithms by comparison with exact results based on posterior model probabilities for the set of 16 models. The explanatory variables were put in studentized form to accommodate the Zellner g-prior used by the MC^3 procedures, which relies on a prior mean of zero for the β coefficients.

The true values for $\alpha, \beta_i, i = 1, \ldots, 4$ were set to values of $\alpha = 5, \beta_1 = 1, \beta_2 = 1$, and $\beta_3 = 0, \beta_4 = 0$. A standard normal random deviate was use for ε in the generating process. The value of ρ was set to 0.6, indicating moderate spatial dependence.

Least-squares estimates as well as maximum likelihood SAR estimates are shown in Table 1. The bias associated with the least-squares estimates appears to be substantial because the two sets of estimates differ widely in magnitude, and the parameter ρ is large and significantly different from zero. The greatest disagreement in the two sets of estimates is with respect to the two spatially lagged explanatory variables, which might be a focus of model comparison and inference. Intuitively, it might be of interest whether housing values and household income levels in nearby neighborhoods contribute to explaining variation in say neighborhood crime rates. We note that the sign of the spatially lagged house value variable is different in the least-squares and SAR regression, and the significance of the spatial lag of household income is different.

As the number of possible models here is $2^4 = 16$, it would have been possible to simply calculate the log-marginal posterior for these 16 models to find posterior model probabilities. Instead, we applied our MC^3 algorithm to least-squares

Table 1 OLS and SAR Model Estimates

Variable	Truth	OLS estimates			SAR estimates		
		Coefficient	t statistic	Probability	Coefficient	t statistic	Probability
Constant	5.0	28.709	4.55	0.000	6.055	0.98	0.323
Income	1.0	1.254	3.16	0.002	0.970	3.15	0.001
hvalue	1.0	0.974	8.99	0.000	0.969	11.64	0.000
$W \cdot hincome$	0.0	2.687	4.22	0.000	0.245	0.41	0.675
$W \cdot hvalue$	0.0	0.242	1.01	0.313	-0.317	-1.42	0.153
ρ					0.637	5.86	0.000

OLS, ordinary least squares; SAR, spatial autoregressive model.

Table 2 OLS BMA Model Selection Information

Variables/models	m1	m2	m3	m4	m5
Income	1	0	0	1	1
hvalue	1	1	1	1	1
$W \cdot hincome$	0	1	1	1	1
$W \cdot hvalue$	1	1	0	1	0
Model probabilities	0.015	0.097	0.097	0.146	0.644

OLS, ordinary least squares; BMA, Bayesian model averaging.

and SARs. A run of 10,000 draws was sufficient to uncover all 16 unique models, requiring 15 and 27 s, respectively, for the OLS and SAR MC^3 procedures.[2]

Information regarding models that exhibited posterior model probabilities >1% is given in Table 2 for the OLS MC^3 procedure and Table 3 for the SAR MC^3 procedure, with the posterior model probabilities shown in the last row of the two tables. These tables use ''1'' and ''0'' indicators for the presence or absence of variables in each of the models presented in the tables.

From the tables, we see that in the case of OLS, the true model was not among the five models with posterior model probabilities >1%. (It was assigned a model probability of 0.1%.) In contrast, the SAR MC^3 procedure identified the true model and assigned it a posterior model probability >50%. This result should not be surprising given that the least-squares estimator is biased and inconsistent in the presence of spatial dependence. The SAR procedure resulted in the two variables, income and housing values used to generate the sample *y* vector, appearing in the four highest probability models that together account for nearly 94% of the posterior probability mass. In contrast, two of the top five least-squares models did not contain both of these variables, and these two models accounted for nearly 20% of the posterior probability mass.

We conclude this discussion by noting that the small sample example based on only 4 candidate variables suggests that the use of least-squares-based MC^3 procedures in the presence of spatial dependence will exert an impact on the model

Table 3 SAR BMA Model Selection Information

Variables/models	m1	m2	m3	m4	m5	m6
Income	0	0	1	1	1	1
hvalue	1	1	1	1	1	1
$W \cdot hincome$	0	0	1	1	0	0
$W \cdot hvalue$	1	0	1	0	1	0
Model probabilities	0.018	0.018	0.051	0.144	0.237	0.506

SAR, spatial autoregressive model; BMA, Bayesian model averaging.

selection inferences. Presumably, this impact is an adverse one as least-squares estimates are either biased and inconsistent for the case of the spatial lag model (SAR), or inefficient in the case of the SEM.

A large sample illustration

An example that models variation in state-to-state migration patterns using origin to destination flows provides a large sample of 2401 observations. Starting with an n by n square matrix of interregional flows from each of the $n = 49$ origin states (including the District of Columbia), to each of the n destination states, we can produce an n^2 by one vector of these flows by stacking the columns of the flow matrix into a variable vector that we designate as y. Without loss of generality, we assume that each of the n columns of the flow matrix represents a different origin and the n rows reflect destinations. The first n elements in the stacked vector y reflect flows from origin 1 to all n destinations. The last n rows of this vector represent flows from origin n to destinations 1 to n. Typically, the diagonal elements of a flow matrix containing flows within a region, for example, from origin 1 to destination 1, origin 2 to destination 2, and so on, will be large relative to the off-diagonal elements representing interregional flows. We set these elements to values of zero to focus our efforts on explaining variation in flows between states. In addition, the vector y was constructed to represent the difference between (logged) state-to-state migration flows during the 1995–2000 period and the flows during the 1985–1990 period. This transformation converts the dependent variable into growth rates of flows over the 10-year period from 1990 to 2000, which produced a relatively normally distributed vector of $49^2 = 2,401$ observations.

A conventional gravity or spatial interaction model based on a nonspatial regression might be used to explain variation in the vector y of origin–destination flows (see Fischer, Scherngell, and Jansenberger 2006; Sen and Smith 1995; Tiefelsdorf 2003). Here, we contrast the nonspatial regression to a spatial regression model of the type suggested by LeSage and Pace (2005). These regression models rely on an n by k matrix of explanatory variables that we label x_d, containing k characteristics for each of the n destinations. Given the format of our vector y, where observations 1 to n reflect flows from origin 1 to all n destinations, this matrix would be repeated n times to produce an n^2 by k matrix of destination characteristics that we represent as X_d for use in the regression. A second matrix containing origin characteristics that we label X_o would be constructed for use in the gravity model. This matrix would repeat the characteristics of the first origin n times to form the first n rows of X_o, the characteristics of the second origin n times for the next n rows of X_o, and so on, resulting in an n^2 by k matrix of origin characteristics. Typically, the distance from each origin to destination is also included as an explanatory variable vector in the gravity model, and perhaps nonlinear terms such as distance-squared. We let D represent an n^2 by 1 vector of these distances from each origin to each destination formed by stacking the columns of the origin–destination distance matrix into a variable vector.[3]

In contrast to the traditional nonspatial gravity model, LeSage and Pace (2005) note that a spatial econometric model of the variation in origin–destination flows would be characterized by: (1) reliance on spatial lags of the dependent variable vector, resulting in a SAR or (2) of the disturbance terms, producing a SEM. Spatial weight matrices represent a convenient and parsimonious way to define the spatial dependence or connectivity relations among observations.

For our applied illustration we use ideas from LeSage and Pace (2005) and rely on a spatial weight matrix consisting of $\tilde{W} = W \otimes I_n + I_n \otimes W$, where W is an n by n spatial weight matrix based on contiguity of the n regions. The n^2 by n^2 matrix $\tilde{W}$ is normalized to have row sums of unity and captures spatial dependence among the origin–destination flows by creating a spatial lag vector $\tilde{W}y$ that averages over neighbors to both origin and destination regions. This leads to a SAR model:

$$y = \rho\tilde{W}y + \alpha\iota_n^2 + X_d\beta_d + X_o\beta_o + D\gamma_1 + D^2\gamma_2 + \varepsilon \qquad (16)$$

In (16), the explanatory variable matrices X_d, X_o represent n^2 by k matrices containing destination and origin characteristics, respectively, and the associated k by 1 parameter vectors are β_d and β_o. The vectors D and D^2 denote the vectorized origin–destination distance matrix and its square with γ_1, γ_2 scalar parameters. As is conventional in SARs, we assume $\varepsilon \sim N(0, \sigma^2 I_{n^2})$.

As candidate explanatory variables, a series of 19 destination and origin specific variables plus distance and distance-squared were used, for a total of 40 explanatory variables. These variables were transformed using logs, so all coefficients should be interpretable as reflecting the elasticity response of the growth rate in state-to-state migration over the period 1990–2000 to changes in each variable. This scaling should help accommodate the prior mean of zero employed in the Zellner g-prior. The distance and distance-squared vectors were also logged. A description of the candidate explanatory variables along with the variable names used to report estimation results is in Table 4.

Runs involving 250,000 draws were used to test for convergence in both OLS and SAR MC^3 results. Two sets of MCMC draws were carried out using randomly selected starting variables in the explanatory variables matrix, resulting in model averaged estimates that were identical to three decimal places for all parameters and in most cases identical to four decimal places.

For the case of least-squares MC^3, the sampling run of 250,000 draws produced 49,246 unique models. The 10 models with the highest posterior model probabilities are shown in Table 5, which takes the same format as Table 2. Variable names are preceded with a symbol ''D-'' or ''O-'' to indicate destination and origin characteristics. The top 10 models accounted for only 44.30% of the probability mass, with 95 models having posterior model probabilities $>0.1\%$, accounting for 83.02% probability mass, 473 models exhibiting model probabilities $>0.01\%$, totaling 95.51% probability mass, 1539 models with probabilities $>0.001\%$ totaling 98.98 probability mass, and 4016 models with probabilities $> 0.0001\%$, accounting for 99.83% of the probability mass.

Table 4 Socioeconomic Demographic Variables

Variable name	Description
Area	The origin–destination state area in square miles
Males	The number of males in 1990
Females	The number of females in 1990
Pcincome	Per capita income in 1990
Young	The # of persons aged 22–29 in 1990
Near retirement	The # of persons aged 60–64 in 1990
Retired	The # of persons aged 65–74 in 1990
Born in state	The # of persons born in the state in 1990
Foreign born	The # of foreign born persons in the state in 1990
Recent immigrants	The # of recent foreign immigrants, during the years 1980–1990
College grads	The # of persons over age 25 with college degrees in 1990
Grad/Profession	The # of persons over age 25 with graduate/professional degrees in 1990
House value	The median house value in 1990
Travel time	Mean travel time to work in 1990
Unemployment	The unemployment rate in 1990
Labor force	The # of persons over 16 years employed in 1990
Median rent	The median rent in 1990
Retirement income	Median retirement income in 1990
Self-employed	The # of persons self-employed in 1990
Distance	The origin–destination state distance
Distance^2	The distance variable squared

To illustrate convergence of the MC^3 sampling process, we note that results from a second run of 250,000 draws based on a random selection of starting variables produced only 47,699 unique models, but the top 10 model probabilities were identical to those reported in Table 5. In addition, there were 96 models having posterior model probabilities $>0.1\%$, accounting for 83.16% probability mass, 473 models exhibiting model probabilities $>0.01\%$, accounting for 95.50% probability mass, 1550 models with probability $>0.001\%$, accounting for 98.97 total probability mass, and 4039 models with probability $>0.0001\%$, accounting for 99.83% of the probability mass. This suggests that the BMA procedure is finding regions of the large model space with posterior support and ignoring regions with low support.

Table 5 shows the variables appearing in the 10 highest posterior probability models. Variables that appear in each model are designated with a ''1'' and those that do not appear with a ''0.'' We find that 13 of the 38 origin–destination variables appear in all of the 10 highest probability models, and these variables are associated with posterior probabilities of inclusion $>65\%$ (probabilities of inclu-

Table 5 Variables Entering the Top 10 Least-Squares Models

Variable name/model	10	9	8	7	6	5	4	3	2	1
D-Area	1	0	0	0	1	1	1	0	1	1
D-Males	1	0	0	0	1	1	1	0	1	1
D-Females	1	1	1	1	1	1	1	1	1	1
D-Pcincome	1	0	0	0	1	1	1	0	1	1
D-Young	0	0	0	0	0	0	0	0	0	0
D-Near retirement	1	1	1	0	1	1	1	1	1	1
D-Retired	0	0	0	1	0	0	0	0	0	0
D-Born in state	1	1	1	1	1	1	1	1	1	1
D-Foreign born	0	0	0	0	0	0	0	0	0	0
D-Recent immigrants	0	1	1	1	0	0	0	1	0	0
D-College grads	0	0	0	0	0	0	0	0	0	0
D-Grad/Profession	0	0	0	0	0	0	0	0	0	0
D-House value	0	0	0	0	0	0	0	0	0	0
D-Travel time	0	0	0	0	0	0	0	0	0	0
D-Unemployment	0	0	0	0	0	0	0	0	0	0
D-Labor force	1	1	1	1	1	1	1	1	1	1
D-Median rent	1	1	1	1	1	1	1	1	1	1
D-Retirement income	1	1	1	1	1	1	1	1	1	1
D-Self-employed	0	0	0	0	0	0	0	0	0	0
O-Area	1	1	1	1	1	1	1	1	1	1
O-Males	0	0	0	0	0	0	0	0	0	0
O-Females	0	0	0	0	0	0	0	0	0	0
O-Pcincome	0	0	0	0	0	0	0	0	0	0
O-Young	0	0	0	0	0	0	0	0	0	0
O-Near retirement	1	1	1	1	1	1	1	1	1	1
O-Retired	1	1	1	1	1	1	1	1	1	1
O-Born in state	0	0	0	0	0	0	0	0	0	0
O-Foreign born	0	0	0	0	0	0	0	0	0	0
O-Recent immigrants	1	1	1	1	1	1	1	1	1	1
O-College grads	0	0	1	0	0	1	0	0	1	0
O-Grad/Profession	1	1	1	1	1	1	1	1	1	1
O-House value	1	1	1	1	1	1	1	1	1	1
O-Travel time	1	1	1	1	1	1	1	1	1	1
O-Unemployment	1	1	1	1	1	1	1	1	1	1
O-Labor force	0	0	0	0	0	0	0	0	0	0
O-Median rent	0	0	0	0	0	0	0	0	0	0
O-Retirement income	1	1	1	0	1	1	0	0	1	0
O-Self-employed	1	0	1	0	0	1	0	0	1	0
Distance	0	0	0	0	0	0	0	0	0	0
Distance^2	1	0	0	0	0	0	1	0	1	0
Model probabilities	0.020	0.020	0.023	0.026	0.027	0.038	0.059	0.069	0.075	0.086

sion are shown in Table 7).[4] One variable ''*D-Near retirement*'' appeared in nine of the 10 highest probability models. After this, we see a decline in variables appearing in only six of the 10 models reported in the table, as well as a decline in the probability of inclusion to below 50%.

A focus of inference for gravity models would be the relative importance of origin versus destination characteristics in explaining variation in state-to-state population migration. Table 5 shows that five of the 13 variables that appear in all of the 10 highest probability models are destination characteristics as well as the one variable that appeared in nine of the 10 top models. This leaves eight of the 13 variables appearing in all 10 top models as origin characteristics, suggesting a slight edge for the case of ''origin push'' as opposed to ''destination pull.'' It will be of interest to contrast this result with those from the SAR.

For the SAR MC^3 procedure, 250,000 draws produced only 5220 unique models, a substantially lower number of models than in the case of least squares. The top 10 models are reported in Table 6, accounting for 76.40% of the total probability mass. Also, in contrast with the least-squares results, we found that 37 models with posterior probabilities $>0.1\%$ accounted for 96.11% of the probability mass and 120 with probabilities $>0.01\%$ accounted for 99.01% of the total probability mass. In summary, the set of high posterior probability models resulting from the spatial model were much more compact than those from least squares. As in the case of least squares, a second run of 250,000 draws produced nearly identical results.

The SAR model MC^3 results are presented in Table 6 in the same format as the least-squares results. In contrast with the least-squares results, only six variables enter all 10 top models, and these are all origin characteristics. The probability for variable inclusion (shown in Table 7) for these six variables was 0.87 or higher. One destination characteristic entered 8 of the 10 top models, ''*D-Grad/Profession*,'' but exhibited a probability of inclusion $<50\%$. The conclusion we would draw here regarding the importance of origin versus destination characteristics is quite different from that reported earlier for least squares. It is interesting that the six origin characteristics that appear in all 10 top SARs also appear in all 10 top least squares models. The spatial autoregressive results appear to exclude a great number of variables from appearing in the model relative to least-squares. There is a theoretical motivation for this type of result. Ignoring the intercept term, we note that least-squares estimates are likely to be biased upward in the face of positive spatial dependence ($\rho > 0$) when the matrix X contains logs of positive values, as shown in the following equation:

$$\hat{\beta}_{SAR} = (X'X)^{-1}X'(I_n - \rho\tilde{W})y$$
$$= \tilde{\beta}_{OLS} - \rho(X'X)^{-1}X'\tilde{W}y \tag{17}$$

Posterior probabilities for each of the 38 origin–destination and distance variables entering the model are shown in Table 7 for both the least-squares and SAR

Table 6 Variables Entering the Top 10 Spatial Autoregressive Models

Variable name/model	10	9	8	7	6	5	4	3	2	1
D-Area	0	0	0	0	0	0	0	0	0	0
D-Males	0	1	1	0	0	1	0	0	0	0
D-Females	1	0	0	0	0	0	1	0	0	0
D-Pcincome	0	0	0	0	0	0	0	0	0	0
D-Young	0	0	0	1	0	0	0	0	0	0
D-Near retirement	1	0	0	0	0	0	0	0	0	0
D-Retired	0	0	1	0	0	0	1	0	0	0
D-Born in state	0	0	1	0	0	1	0	1	0	0
D-Foreign born	0	0	0	0	0	0	0	0	0	0
D-Recent immigrants	0	0	0	0	0	0	0	0	0	0
D-College grads	0	0	0	0	0	0	0	0	0	0
D-Grad/Profession	1	1	0	1	1	0	1	1	1	1
D-House value	0	0	0	0	0	0	0	0	0	0
D-Travel time	0	1	0	1	1	0	0	1	1	1
D-Unemployment	0	1	0	1	1	0	0	1	1	1
D-Labor force	0	0	1	0	0	1	0	0	0	0
D-Median rent	0	0	0	0	0	0	0	0	0	0
D-Retirement income	0	0	0	0	0	0	0	0	0	0
D-Self-employed	0	0	0	0	0	0	0	0	0	0
O-Area	1	1	1	1	1	1	1	1	1	1
O-Males	0	0	0	0	0	0	0	0	0	0
O-Females	0	0	0	0	0	0	0	0	0	0
O-Pcincome	0	0	0	0	0	0	0	0	0	0
O-Young	0	0	0	0	0	0	0	0	0	0
O-Near retirement	1	1	1	1	1	1	1	1	1	1
O-Retired	1	1	1	1	1	1	1	1	1	1
O-Born in state	0	0	0	0	0	0	0	0	0	0
O-Foreign born	0	0	0	0	0	0	0	0	0	0
O-Recent immigrants	1	1	1	1	1	1	1	1	1	1
O-College grads	0	0	0	0	0	0	0	0	0	1
O-Grad/Profession	0	0	0	0	0	0	0	0	1	0
O-House value	1	1	1	1	1	1	1	1	1	1
O-Travel time	0	0	0	0	0	0	0	0	0	0
O-Unemployment	1	1	1	1	1	1	1	1	1	1
O-Labor force	0	0	0	0	0	0	0	0	0	0
O-Median rent	0	0	0	0	0	0	0	0	0	0
O-Retirement income	0	0	0	0	0	0	0	0	0	0
O-Self-employed	0	0	0	0	0	0	0	0	0	0
Distance	0	0	0	0	1	0	0	0	0	0
Distance^2	0	0	0	0	0	0	0	0	0	0
Model probabilities	0.023	0.025	0.026	0.029	0.035	0.040	0.045	0.063	0.112	0.366

Table 7 Posterior Probabilities for Variables Entering the Model

Variable name	OLS MC^3	SAR MC^3	Difference
D-Area	0.3134	0.0219	0.2915
D-Males	0.2146	0.0231	0.1915
D-Females	0.7947	0.0206	0.7741
D-Pcincome	0.1799	0.0225	0.1574
D-Young	0.0657	0.0191	0.0466
D-Near retirement	0.5013	0.0224	0.4789
D-Retired	0.3745	0.0235	0.3510
D-Born in state	0.9406	0.0231	0.9175
D-Foreign born	0.0499	0.0159	0.0340
D-Recent immigrants	0.4531	0.0176	0.4355
D-College grads	0.0662	0.0205	0.0457
D-Grad/Profession	0.0601	0.4790	− 0.4189
D-House value	0.0838	0.0246	0.0592
D-Travel time	0.0730	0.0990	− 0.0260
D-Unemployment	0.2562	0.1015	0.1547
D-Labor force	0.9382	0.4517	0.4865
D-Median rent	0.7327	0.0213	0.7114
D-Retirement income	0.7777	0.0228	0.7549
D-Self-employed	0.0403	0.0214	0.0189
O-Area	0.8510	0.9412	− 0.0902
O-Males	0.1094	0.0280	0.0814
O-Females	0.0794	0.0291	0.0503
O-Pcincome	0.1511	0.0275	0.1236
O-Young	0.0727	0.0205	0.0522
O-Near retirement	0.8562	0.9089	− 0.0527
O-Retired	0.8761	0.8917	− 0.0156
O-Born in state	0.1198	0.0221	0.0977
O-Foreign born	0.0766	0.0211	0.0555
O-Recent immigrants	0.8555	0.8791	− 0.0236
O-College grads	0.2079	0.0233	0.1846
O-Grad/Profession	0.8536	0.0220	0.8316
O-House value	0.9403	0.8907	0.0496
O-Travel time	0.6508	0.0214	0.6294
O-Unemployment	0.8661	0.8915	− 0.0254
O-Labor force	0.1164	0.0224	0.0940
O-Median rent	0.2189	0.0218	0.1971
O-Retirement income	0.3633	0.0214	0.3419
O-Self-employed	0.4294	0.0208	0.4086
Distance	0.1238	0.0204	0.1034
Distance^2	0.3943	0.0207	0.3736

OLS, ordinary least squares; SAR, spatial autoregressive model.

MC^3 procedures. The last column in the table shows the difference between the OLS and SAR model results. We see 6 of 40 cases where the differences are >50%, and another 8 of 40 cases where these differences are between 0.29 and 0.48, pointing to a number of cases where the inclusion probabilities from the least-squares procedure are higher than the spatial model. In contrast, we see seven negative differences, with one equal to -0.4189, and the remainder < -0.10, reflecting a small number of cases with higher variable inclusion rates for the SAR. The average number of variables appearing in the top 10 least-squares models was 17.8 and 9.7 for the SARs.

Estimates for the variables would allow an examination of the elasticity response and direction of impact on population migration associated with the variables that exhibit high probabilities of inclusion. Table 8 reports model-averaged SAR estimates based on averaging over the 120 models with posterior probabilities >0.01%, and OLS estimates based on averaging over 473 models exhibiting model probabilities >0.01%.[5] Bayesian MCMC estimates for the OLS and SAR model implemented the Zellner's *g*-prior, diffuse intercept and noise variance priors, and the *Beta* (1.01, 1.01) prior for ρ were used to produce 2000 retained draws. These draws were weighted by the posterior model probabilities and used to construct a posterior mean as well as 5% and 95% highest posterior density intervals (HPDI) that are reported in Table 8.

For interpretation purposes, the mean growth rate in interstate migration over the 10-year period was 0.0465, or less than one-half percent per year, with the 5 percentile being -0.4731 or around negative 5% per year and the 95 percentile being 0.5868, or about 6% per year. An estimate for β of 0.10 suggests that a 10% change in the explanatory variable would give rise to a 1% change in the growth rate over the 10-year period, and we focus on coefficient estimates that are >0.10 in absolute value when analyzing the model-averaged estimates.

For the case of the SAR model estimates, there were five origin characteristics that exerted impacts greater than 0.10 (in absolute value) on migration flows, *O-Unemployment*, *O-Retired*, *O-Near retirement*, *O-House value*, and *O-Recent immigrants*. Of these, the unemployment rate, retired persons, and house value were positive, while persons near retirement and recent immigrants were negative. The origin unemployment rate had the largest impact on out-migration flows, with the estimate suggesting that a state with a 1% higher unemployment rate would exhibit a 2% higher growth rate in out-migration over the 10-year period. Retired persons and those aged near retirement had the second largest impact, both exhibiting elasticities around unity, but with opposite signs. Intuitively, these signs seem correct. Higher house values in the origin state lead to out-migration flows and a higher number of recent immigrants from abroad lead to lower out-migration, which seem intuitively plausible. There were only three destination characteristics that exerted an impact >0.10 on migration growth rates, *D-Unemployment*, *D-Graduate/Professional*, and *D-Travel time*, and all were <0.5639 in magnitude. The travel time and unemployment rates had a positive impact on in-migration

Table 8 Model Averaged Estimates

Variable name	SAR estimates			OLS estimates		
	5% HPDI	Mean	95% HPDI	5% HPDI	Mean	95% HPDI
Constant	− 13.2243	− 12.4719	− 11.6694	− 15.1984	− 14.6921	− 14.1894
D-Area	− 0.0000	0.0001	0.0003	0.0452	0.0496	0.0537
D-Males	0.0101	0.0129	0.0156	− 1.4916	− 1.3363	− 1.1766
D-Females	0.0111	0.0159	0.0207	2.2110	2.3858	2.5532
D-Pcincome	0.0000	0.0000	0.0000	0.2743	0.3046	0.3348
D-Young	− 0.0005	− 0.0002	0.0001	0.0000	0.0000	0.0000
D-Near retirement	− 0.0090	− 0.0065	− 0.0042	− 0.7743	− 0.7439	− 0.7119
D-Retired	− 0.0209	− 0.0166	− 0.0125	− 0.0833	− 0.0790	− 0.0745
D-Born in state	− 0.0095	− 0.0075	− 0.0054	− 0.3748	− 0.3611	− 0.3475
D-Foreign born	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
D-Recent immigrants	0.0000	0.0000	0.0000	0.1013	0.1124	0.1231
D-College grads	− 0.0004	− 0.0001	0.0001	0.0003	0.0005	0.0006
D-Grad/Profession	− 0.1719	− 0.1589	− 0.1470	0.0000	0.0000	0.0000
D-House value	0.0045	0.0061	0.0077	0.0014	0.0017	0.0021
D-Travel time	0.1462	0.1744	0.2018	0.0022	0.0027	0.0033
D-Unemployment	0.5639	0.6779	0.7991	0.0575	0.0620	0.0665
D-Labor force	− 0.0331	− 0.0301	− 0.0270	− 0.9079	− 0.8786	− 0.8505
D-Median rent	0.0000	0.0000	0.0000	− 0.4110	− 0.3890	− 0.3671
D-Retirement income	− 0.0005	− 0.0004	− 0.0003	− 0.2321	− 0.2204	− 0.2084
D-Self-employed	− 0.0001	0.0001	0.0004	0.0000	0.0000	0.0000
O-Area	0.0446	0.0493	0.0541	0.0930	0.0954	0.0979
O-Males	− 0.0046	− 0.0015	0.0014	0.0000	0.0000	0.0000
O-Females	− 0.0006	− 0.0004	− 0.0003	0.0000	0.0000	0.0000
O-Pcincome	− 0.0014	0.0004	0.0021	− 0.0007	− 0.0005	− 0.0002
O-Young	− 0.0016	− 0.0002	0.0012	0.0000	0.0000	0.0000
O-Near retirement	− 1.0307	− 0.9335	− 0.8415	− 1.7331	− 1.6795	− 1.6257
O-Retired	1.0556	1.1530	1.2505	1.9685	2.0223	2.0754
O-Born in state	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
O-Foreign born	− 0.0002	0.0000	0.0003	0.0000	0.0000	0.0000
O-Recent immigrants	− 0.3757	− 0.3348	− 0.2930	− 0.5383	− 0.5215	− 0.5047
O-College grads	0.0230	0.0562	0.0919	0.0495	0.0598	0.0700
O-Grad/Profession	0.0014	0.0061	0.0109	0.2039	0.2137	0.2233
O-House value	0.6192	0.6779	0.7379	1.0702	1.0993	1.1294
O-Travel time	0.0000	0.0000	0.0000	− 0.1930	− 0.1776	− 0.1620
O-Unemployment	1.9197	2.0490	2.1771	2.3486	2.4341	2.5200
O-Labor force	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
O-Median rent	− 0.0009	0.0003	0.0016	0.0122	0.0150	0.0175
O-Retirement income	0.0000	0.0000	0.0000	− 0.0999	− 0.0915	− 0.0834
O-Self-employed	0.0000	0.0000	0.0000	− 0.0286	− 0.0260	− 0.0234
Distance	− 0.0000	0.0002	0.0004	− 0.0011	− 0.0008	− 0.0006
Distance^2	0.0000	0.0000	0.0000	0.0007	0.0008	0.0009

OLS, ordinary least squares; SAR, spatial autoregressive model; HPDI, highest posterior density intervals.

flows to the destination, which seems implausible, whereas persons with graduate/ professional degrees had a negative impact, a plausible result.

Taken together, the SAR model-averaged estimates suggest that origin characteristics are relatively more important than destination characteristics, leading to a ''push'' interpretation of migration flows.

The OLS model-averaged estimates were different in that they pointed to nine destination characteristics having coefficients >0.10 (in absolute) and seven origin characteristics meeting this criterion. As in the case of the SAR model-averaged estimates, the *O-Unemployment* had the largest impact with a coefficient of 2.4341. As indicated earlier, this is likely to exhibit an upward bias relative to the SAR model estimate. Similar upward biases can be found in OLS estimates for *O-Near retirement* and *O-Retired* and *O-House value*. The second largest impact was *D-Females* which taken together with the large negative impact for *D-Males*, suggests a destination-state population size effect, as males plus females equal population. Interestingly, in the SAR that contains a spatial lag variable, this population size effect is not noticeable.

In total, the OLS-model averaged estimates suggest a different inference regarding the relative importance of origin versus destination characteristics on migration flows than that described above for the SAR model. We note that the model-averaged posterior mean estimate for the parameter ρ was 0.6077, with a 5% and 95% HPDI of 0.5900 and 0.6252, respectively, suggesting a bias in the least-squares results.

Conclusion

It is often the case that sample data exhibit spatial dependence requiring use of spatial regression models of the type considered here. Spatial dependence can lead to least-squares estimates that are biased and inconsistent, and at best least-squares will produce inefficient estimates. This invalidates use of conventional least-squares MC^3 and BMA techniques. We have developed the theory as well as computationally efficient algorithms for implementing MC^3 and BMA techniques for two important spatial regression models. Expressions for the log-marginal posterior based on Zellner's g-prior were developed and found to be considerably more complex than in the case of least squares. Implementation of MC^3 techniques for the spatial regression models requires repeated evaluation of the log-marginal posterior in the context of Metropolis–Hastings sampling. Numerical integration over the spatial dependence parameter as well as calculation of the log-determinant of an n by n matrix are required to evaluate the log-marginal posterior. We provide computational details for efficient algorithms that allow this to be done in a rapid fashion, making spatial regression model MC^3 competitive with conventional least-squares MC^3 algorithms. A set of public domain MATLAB functions that implement the methods described here are available at www.spatial-econometrics.com for public use.

An applied illustration using a SAR approach to model state-to-state population migration flows over the period 1990–2000 demonstrated that different estimates and inferences result from using an OLS and SAR MC^3 methodology in conjunction with model averaging. The results differed regarding which variables were important in explaining variation in population migration flows as well as the model-averaged estimates of the relative impact of the various variables.

Acknowledgement

Appendix A

There are four separate terms involved in the univariate integration problem over the range of support for the parameter ρ in the SAR model. In the context of the MC^3 described in the third section, there is a need for a computationally fast scheme for the univariate integration. This must be carried out on every pass through the MCMC sampler that occurs thousands of times.

The four terms in (11) for the SAR model that vary with ρ are shown in (A1) as T_1, T_2, T_3, and T_4.

$$
\begin{aligned}
T_1(\rho) &= |I_n - \rho W| \\
T_2(\rho) &= [(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n]'[(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n] \\
T_3(\rho) &= (\frac{g}{1+g})\hat{\beta}(\rho)X'X\hat{\beta}(\rho) \\
T_4(\rho) &= \frac{1}{Beta(\alpha_0, \alpha_0)}\frac{(1+\rho)^{\alpha_0-1}(1-\rho)^{\alpha_0-1}}{2^{2\alpha_0-1}}
\end{aligned}
\tag{A1}
$$

A log transformation can be applied to all terms $T_1, \ldots, T_4$, allowing us to rely on computationally fast methods presented in Pace and Barry (1997) and Barry and Pace (1999) to compute the log-determinant in T_1, that is, $\ln (\det (I_n - \rho W))$. One of the earlier computationally efficient approaches to solving for estimates in a model involving a sample of 3107 observations was proposed by Pace and Barry (1997). They suggested using direct sparse matrix algorithms such as the Cholesky or LU decompositions to compute the log determinant. In addition to storage savings, sparse matrices result in lower operation counts as well, speeding computations.

Pace and Barry (1997) also suggest a vectorization of the terms in T_1 and T_2 of our problem that they found useful for maximum likelihood estimation of the SAR model. This involves constructing log-determinant values over a grid of q values of ρ, which is central to our task of integration for the terms $T_1(\rho)$ and $T_2(\rho)$. In the SAR and SEM models, support for ρ must lie in the interval $[\mu_{min}^{-1}, 1]$, but typical applied

work simply relies on a restriction of ρ to the $(-1, 1)$ or $(0, 1)$ interval to avoid the need to compute eigenvalues.

Turning attention to the term $T_2(\rho)$, Pace and Barry (1997) write the term, $[(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n]'[(I_n - \rho W)y - X\hat{\beta}(\rho) - \hat{\alpha}\iota_n]$ as a vector in q values of ρ. For our problem, we have the expression shown in the following equation:

$$T_2(\rho_i) = e(\rho_i)'e(\rho_i), \forall i = 1, \ldots, q \tag{A2}$$

with

$$e(\rho_i) = e_o - \rho_i e_d$$

$$e_o = y - X\beta_o - \alpha_o \iota_n$$

$$e_d = Wy - X\beta_d - \alpha_d \iota_n$$

$$\beta_o = (X'X)^{-1}X'y$$

$$\beta_d = (X'X)^{-1}X'Wy$$

$$\alpha_o = \bar{y}$$

$$\alpha_d = \overline{Wy} \tag{A3}$$

The term T_3 can be vectorized using a loop over ρ_i values along with the expression $\hat{\beta}(\rho) = \beta_o - \rho_i\beta_d$. This presents the most problems for a matrix programming language such as MATLAB, as a loop must be used to evaluate the quadratic form in T_3 for each value of ρ_i.

Finally, the term $T_4(\rho)$ representing the prior on the parameter ρ is simple to compute over a grid of q values for ρ, and transform to logs.

One important point to note is that we need not estimate the parameters $\eta = (\alpha, \beta, \sigma, \rho)$ of the model to carry out numerical integration leading to posterior model probabilities. Intuitively, we have analytically integrated the parameters α, β, and σ out of the problem, leaving only a univariate integral in ρ. Given any sample data y, X along with a spatial weight matrix W, we can rely on the Pace and Barry (1997) vectorization scheme applied to our task. This involves evaluating the log-marginal density terms $T_1, \ldots, T_4$ over a fine grid of q values for ρ ranging over the interval, $[\mu_{min}^{-1}, 1]$, or $[-1, 1]$. Given a matrix of vectorized log-marginal posteriors, integration can be accomplished using Simpson's rule.

Further computational savings can be achieved by noting that the grid can be rough, say based on 0.01 increments in ρ, which speeds the direct sparse matrix approach of Pace and Barry (1997) or Barry and Pace (1999) computations. Spline interpolation can then be used to produce a much finer grid very quickly, as the log determinant is typically quite well behaved for reasonably large spatial samples in excess of 250 observations.

Another important point concerns scaling necessary to carry out the numerical integration that is carried out on the antilog of the log-marginal posterior density. Our approach allows one to evaluate log-marginal posteriors for each model under consideration and store these as vectors ranging over the grid of ρ values. Scaling can then be simply solved by finding the maximum of these vectors placed as columns in a matrix, that is, the maximum from all columns in the matrix. This maximum is then subtracted from all elements in the matrix of log-marginals, producing a value of zero as the largest element, so that the anti-log is unity. This approach to scaling provides an eloquent and universal solution that requires no user intervention and works for all problems.

Numerical integration over λ for the SEM model can use the same vectorization for the terms T_1 and T_4 as in the case of the SAR model. For this model, we need to construct vectors for $y^* = y - \lambda_i W y$ and $X^* = X - \lambda_i W X$, that are used to calculate a vectorized version of $T_2 = (1/(1+g))[y^* - X^*\hat{\beta} - \hat{\alpha}\iota]'[y^* - X^*\hat{\beta} - \hat{\alpha}\iota]$.

The term T_3 expression for the SEM model which contains an intercept term takes the form $T_3 = (g/(1+g))(y^* - \hat{\alpha}\iota)'(y^* - \hat{\alpha}\iota)$, which can be vectorized over the λ_i values in y^*.

Notes

1 There are also other models in the family elaborated in Anselin (1988) that are possible choices.
2 MATLAB version 7 software was used in conjunction with a Pentium III M laptop computer.
3 The diagonal elements of the distance matrix containing distances from origin 1 to destination 1, origin 2 to destination 2, and so on, will be zero.
4 Fernández, Ley, and Steel (2001a, b) provide details on calculations of probabilities for inclusion of individual variables in the models.
5 It is common practice to reweight the posterior model probabilities so that they sum to unity and use these normalized weights to produce model-averaged estimates. This was done here.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, Germany: Kluwer Academic Publishers.

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

Denison, D. G. T., B. K. Mallick, and A. F. M. Smith. (1998). ''Automatic Bayesian Curve Fitting.'' *Journal of the Royal Statistical Society, Series B* 60, 333–50.

Fernández, C., E. Ley, and M. F. J. Steel. (2001a). ''Model Uncertainty in Cross-Country Growth Regressions.'' *Journal of Applied Econometrics* 16, 563–76.

Fernández, C., E. Ley, and M. F. J. Steel. (2001b). ''Benchmark Priors for Bayesian Model Averaging.'' *Journal of Econometrics* 100, 381–427.

Fischer, M. M., T. Scherngell, and E. Jansenberger. (2006). ''The Geography of Knowledge Spillovers between High-Technology Firms in Europe Evidence from a Spatial Interaction Modelling Perspective.'' *Geographical Analysis* 38, 288–309.

Florax, R. J. G. M., H. Folmer, and S. J. Rey. (2003). ''Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology.'' *Regional Science and Urban Economics* 33, 557–79.

Hepple, L. W. (1995a). ''Bayesian Techniques in Spatial and Network Econometrics: 1. Model Comparison and Posterior Odds.'' *Environment and Planning A* 27, 447–69.

Hepple, L. W. (1995b). ''Bayesian Techniques in Spatial and Network Econometrics: 2. Computational Methods and Algorithms.'' *Environment and Planning A* 27, 615–44.

LeSage, J. P., and R. K. Pace. (2004). ''Introduction to Spatial and Spatiotemporal Econometrics.'' In *Advances in Econometrics: Vol. 18. Spatial and Spatiotemporal Econometrics*, edited by J. P. LeSage and R. K. Pace. Oxford, UK: Elsevier Ltd.

LeSage, J. P., and R. K. Pace. (2005). ''Spatial Econometric Modeling of Origin–Destination Flows.'' Paper presented at the Regional Science Association International Meetings, Las Vegas, NV, November 2005.

Lindley, D. V. (1957). ''A Statistical Paradox.'' *Biometrika* 44, 187–92.

Madigan, D., and J. York. (1995). ''Bayesian Graphical Models for Discrete Data.'' *International Statistical Review* 63, 215–32.

Ord, J. K. (1975). ''Estimation Methods for Models of Spatial Interaction.'' *Journal of the American Statistical Association* 70, 120–26.

Pace, R. K., and R. Barry. (1997). ''Sparse Spatial Autoregressions.'' *Statistics and Probability Letters* 33, 291–97.

Raftery, A. E., D. Madigan, and J. A. Hoeting. (1997). ''Bayesian Model Averaging for Linear Regression Models.'' *Journal of the American Statistical Association* 92, 179–91.

Richardson, S., and P. J. Green. (1997). ''On Bayesian Analysis of Mixtures with an Unknown Number of Components.'' *Journal of the Royal Statistical Society B* 59, 731–92.

Sen, A., and T. E. Smith. (1995). *Gravity Models of Spatial Interaction Behavior.* Heidelberg, Germany: Springer-Verlag.

Sun, D., R. K. Tsutakawa, and P. L. Speckman. (1999). ''Posterior Distribution of Hierarchical Models Using Car(1) Distributions.'' *Biometrika* 86, 341–50.

Tiefelsdorf, M. (2003). ''Misspecifications in Interaction Model Distance Decay Relations: A Spatial Structure Effect.'' *Journal of Geographical Systems* 5, 25–50.

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

Zellner, A. (1986). ''On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions.'' In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, edited by P. Goel and A. Zellner, 233–34. Amsterdam: North-Holland/Elsevier.