

# Flexible dependence modeling using convex combinations of different types of connectivity structures

Nicolas Debarsy  
CNRS- LEM UMR 9221  
Université de Lille 1  
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 explore alternative approaches for constructing convex combinations of different types of dependence between observations. Pace and LeSage (2002) as well as Hazir et al. (2016) use convex combinations of different connectivity matrices to form a single weight matrix that can be used in conventional spatial regression estimation and inference. An example for the case of two weight matrices,  $W_1, W_2$  reflecting different types of dependence between a cross-section of regions, firms, individuals etc., located in space would be:  $W_c = \gamma_1 W_1 + (1 - \gamma_1) W_2$ ,  $0 \leq \gamma_1 \leq 1$ . The matrix  $W_c$  reflects a convex combination of the two weight matrices, with the scalar parameter  $\gamma_1$  indicating the relative importance assigned to each type of dependence. We explore issues that arise in producing estimates and inferences from these more general cross-sectional regression relationships in a Bayesian framework. We propose two procedures to estimate such models and assess their finite sample properties through Monte Carlo experiments. We illustrate our methodology in an application to CEO salaries for a sample of nursing homes located in Texas. Two types of weights are considered, one reflecting spatial proximity of nursing homes and the other peer group proximity, which arises from the salary benchmarking literature.

**KEYWORDS:** Spatial econometrics, Connectivity matrix, Salary benchmarking models, Markov Chain Monte Carlo estimation, Bayesian model probabilities, convex combination.

# 1 Introduction

Spatial regression models typically rely on spatial proximity to specify weight matrices, for example first-order neighbors (those with borders in common) or some number (say  $m$ ) of nearest neighboring regions, or points (e.g., firms, consumers, houses) located in space. This approach has two advantages: 1) geographical location of observations is objective and easy to determine, and 2) weight matrices based on geographical space can be viewed as fixed over time and in most cases exogenous.<sup>1</sup> There has been a great deal of criticism of weight matrices based solely on spatial location of observations, (e.g., Corrado and Fingleton, 2012). This criticism in part derives from application of spatial regression models to broader contexts involving interregional flows of: goods (e.g., trade), population (e.g., migration), knowledge (e.g., patent citations); student peer groups, social networks, etc., where geographical location of observations does not seem intuitively or theoretically appealing.

There are a limited number of studies where weight matrices reflecting connectivity of observations have been motivated by underlying theoretical considerations. For example Behrens et al. (2012) derive a quantity-based structural gravity equation system where both trade flows and error terms are cross-sectionally correlated based on population shares of regions in the sample, and Koch and LeSage (2015) show that the multilateral resistance concept from trade theory (Anderson and van Wincoop, 2003, 2004) can be viewed as a simultaneous autoregressive dependence structure involving gross domestic product shares of the sample regions as well as other types of generalized distance factors.

One response to dissatisfaction regarding use of connectivity structures based solely on spatial location has been the introduction of simultaneous autoregressive models that rely on more than one weight matrix (see Lacombe, 2004; Badinger and Egger, 2011; Lee and Liu, 2010; Elhorst et al., 2011; Liu et al., 2014). In these models, different weight matrices are introduced in an effort to capture different types of cross-sectional dependence within the same spatial autoregressive specification. Specifically, multiple spatial lags of the dependent variable vector  $y$  are introduced as right-hand side variables in an effort to extend conventional spatial autoregressive models to include what have been labeled ‘higher-order’ terms, as shown in (1), where  $L$  of such terms are introduced.

---

<sup>1</sup>There are cases where location of firms or consumers in space could result from an endogenous sorting process or some other underlying economic mechanism.

$$y = \left( \sum_{\ell=1}^L \rho_{\ell} W_{\ell} \right) y + X\beta + \varepsilon \quad (1)$$

In (1),  $y$  is an  $n \times 1$  vector of dependent variable outcomes,  $X$  is an exogenous  $n \times k$  explanatory variables matrix, with  $\beta$  the associated  $k \times 1$  parameters,  $\rho_{\ell}$ ,  $\ell = 1, \dots, L$  are scalar dependence parameters measuring the strength of each type of dependence modeled by the  $n \times n$  connectivity matrices  $W_{\ell}$ ,  $\ell = 1, \dots, L$ . The  $n \times 1$  disturbance vector is assumed to have independent error terms with zero mean and constant scalar variance ( $\sigma^2$ ) across all  $n$  observations. The specification (1) has been labeled a *higher-order* spatial autoregressive model in the literature as it includes more than one spatial lag of the dependent variable ( $Wy$ ) (Lee and Liu, 2010; Lee and Yu, 2014).

One example of a higher-order specification is Lacombe (2004), who explored a county-level cross-sectional spatial relationship involving program participation of state residents, where a spatial matrix  $W_1$  is used to identify nearby counties located within the same state and a second spatial weight matrix  $W_2$  captures the influence of nearby counties located in neighboring states. Liu et al. (2014) in a model of social interaction that explores peer effects rely on one weight matrix to capture local-average (social norm) influences of peers and a second weight matrix for influences reflected by local-aggregate peer effects (social multiplier). However, LeSage and Pace (2011) point to a number of estimation and interpretive issues that arise for models of the type in (1), and Elhorst et al. (2011) point out complications that arise regarding the parameter space for the dependence parameters  $\rho_{\ell}$ ,  $\ell = 1, \dots, L$ .

Another thread in the literature is to simply replace the spatial weight matrix with more appropriate types of connectivity structures, for example weight matrices based on friendship ties. Patacchini and Zenou (2012) analyze the role played by teenagers conformity to their peers' behavior in producing juvenile crime outcomes in a social networking application. In the area of international finance, connectivity matrices may reflect real transmission channels for risk across countries, which might arise from trade or financial ties between countries. Alternatively, information transmission channels for risk might be reflected by financial market information that captures perceptions of market participants regarding own- and other-country risks (see Debarsy et al., 2016, and references therein). This suggests weight matrices based on trade, financial flows, or socio-economic similarities

between countries.

A related literature is on methods for assessing different weight structures for their consistency with a specific economic relationship and set of sample data. Since models based on alternative weight matrices are likely to be non-nested, one approach in this literature uses the non-nested  $J$  test developed by Davidson and MacKinnon (1981) extended to spatial regression models by Anselin (1988). The power of alternative predictions for a host of spatial regression model specifications are explored in Kelejian (2008) and Kelejian and Piras (2011). Burridge and Fingleton (2010) and Burridge (2012) propose respectively to rely on bootstrap procedures for inference on the  $J$  test and to use maximum likelihood estimation rather than instrumental variables in the first step of the  $J$  test computation, within the Kelejian (2008) framework. Liu et al. (2014) propose an extension of the Kelejian (2008)  $J$  test to differentiate between the local-aggregate and the local-average endogenous peer effects in an econometric network model with network fixed effects. Debarsy and Ertur (2016) build on the  $J$  tests of Kelejian (2008) and Kelejian and Piras (2011) to allow for heteroskedasticity in a spatial autoregressive specification and further propose a procedure based on Hagemann (2012) to circumvent the decision problem inherent to non-nested models tests (the decision problem arises when non-nested tests do not lead to a clear choice between competing models). Alternatively, Jin and Lee (2013) consider a spatial model extension of the Cox test (Cox, 1961, 1962) for the case of non-nested models. Finally, in the context of determining the most relevant geographically based spatial weight matrix, Geniaux and Martinetti (2017) suggest to use different distance kernels with a single parameter  $h$  (representing the bandwidth or the number of neighbors, depending on the kernel). Identification of the matrix  $W$  is then based on a moment estimator that tries to minimize the residual sum of squares of the model estimation with respect to  $W(h)$ .

A Bayesian alternative to non-nested model tests is proposed by LeSage and Pace (2009) in order to select the most appropriate spatial weight matrix. In contrast to the  $J$  tests that rely on specific model estimates and their associated predictions, the Bayesian approach to model comparison integrates over all model parameters to calculate the log-marginal likelihood and associated model probabilities. This approach makes inference regarding the best weight matrix unconditional on any particular set of estimates. Note that parameter estimates based on models that rely on the wrong weight matrix will be biased, making it desirable to draw model comparison conclusions that are unconditional on the parameter

estimates. LeSage (2014, 2015), extends this approach to simultaneously calculate log-marginal likelihoods and associated model probabilities for both cross-sectional and panel data model specifications and weight matrices. Again, inferences drawn based on posterior model probabilities are unconditional on parameter estimates from the host of alternative models considered.

Our contribution to the spatial econometric literature regarding alternative weight matrices is to pursue an approach considered by Pace and LeSage (2002) as well as in Hazir et al. (2016), that relies on convex combinations of different connectivity matrices to form a single weight matrix. An advantage of this approach is that the resulting weight matrix can be used in conventional spatial regression methods to produce estimates and inference. This means that existing software algorithms that are widely available for estimation and interpretation of spatial regression models can be used. This approach avoids many of the issues raised in LeSage and Pace (2011) regarding estimation and interpretation of higher-order models that include spatial lags involving multiple different  $W$  matrices.

This convex combination approach proposes using  $W_c = \sum_{\ell=1}^L \gamma_\ell W_\ell$ , with  $0 \leq \gamma_\ell \leq 1$ ,  $\ell = 1, \dots, L$  and  $\sum_{\ell=1}^L \gamma_\ell = 1$ , in a standard spatial econometrics specification. The matrix  $W_c$  reflects cross-section dependence specified using a convex combination of  $L$  different types of connectivity between observations.<sup>2</sup> The scalar parameters  $\gamma_\ell$  indicate the relative importance assigned to each type of dependence in the cross-sectional dependence scheme. The two sets of constraints imposed reflect the fact that this approach relies on a convex combination. When each  $W_\ell, \ell = 1, \dots, L$ , is row-normalized, then  $W_c$  obeys the conventional row-normalization, which allows use of conventional spatial regression model specifications and estimation methods.<sup>3</sup>

In Section 2, we explore two alternative estimation strategies for determining estimates of  $\gamma_\ell, \ell = 1, \dots, L$  in this convex combination approach, one that calculates Bayesian posterior model probabilities for a discrete set of  $M$  models based on a grid of values for  $\gamma_\ell \in [0, 1], \ell = 1, \dots, L$ , and another based on Bayesian MCMC estimation of the posterior

---

<sup>2</sup>In this paper, we do not address the potential endogeneity issue that may arise when weight matrices are not based on geographic proximity. The matrices entering the convex combination are thus assumed to be exogenous. For standard SAR cross-section models, Qu and Lee (2015) develop an estimator robust to endogeneity of the connectivity structure. However, in our context, we leave this question for further research.

<sup>3</sup>Alternative types of normalization for connectivity matrices  $W_\ell$  are possible (see Kelejian and Prucha, 2010). However, with the exception of special cases, normalization of each connectivity matrix  $W_\ell$  by one of the matrix norms proposed in Kelejian and Prucha (2010) does not result in a normalized convex combination matrix  $W_c$ .

distributions for the parameters  $\gamma_\ell$ . Both the posterior distributions and model probabilities provide a formal basis for inference regarding the relative importance assigned to each of the connectivity structures  $W_\ell$ .

Guidance for successful use of these models by practitioners is provided in Section 3 of the paper. Section 4 presents results from Monte Carlo experiments, where we show that the Bayesian MCMC approach to estimating  $\gamma_\ell$ ,  $\ell = 1, \dots, L$  provides superior coverage intervals for the underlying model parameters and more reliable scalar summary measures of direct and indirect effects used to draw inferences regarding the impact of a change in the explanatory variables on the outcome variable. An applied illustration is the subject of Section 5.

In the remainder of this paper we consider a host of spatial specifications that can be used in conjunction with the convex combination of weights that we have labeled  $W_c$ . These models are shown in expressions (2) to (6). The only difference with respect to traditional spatial econometrics specifications comes from the use of a connectivity matrix  $W_c$  which represents a global measure of cross-sectional dependence, combining different specifications regarding the nature of interactions between observations. In traditional spatial models used in the literature, the connectivity matrix models a single type of cross-sectional dependence that has been labeled spatial dependence.

$$\text{SAR} : y = \rho W_c y + X\beta + \varepsilon \quad (2)$$

$$\text{SLX} : y = X\beta + W_c X\theta + \varepsilon \quad (3)$$

$$\text{SDM} : y = \rho W_c y + X\beta + W_c X\theta + \varepsilon \quad (4)$$

$$\text{SDEM} : y = X\beta + W_c X\theta + (I_N - \lambda W_c)^{-1} u. \quad (5)$$

$$\varepsilon, u \sim N(0, \sigma^2 I_{N^2})$$

$$W_c = \sum_{\ell=1}^L \gamma_\ell W_\ell \quad (6)$$

$$\text{s.t. } 0 \leq \gamma_\ell \leq 1, \ell = 1, \dots, L \text{ and } \sum_{\ell=1}^L \gamma_\ell = 1$$

## 2 Convex combinations of connectivity matrices

Pace and LeSage (2002) as well as Hazir et al. (2016) use convex combinations of different weight matrices to form a single connectivity matrix that can be used in conventional spatial regression estimation. One advantage of this approach is that estimation procedures for traditional spatial regression models can be used to produce estimates and inference for models (2) to (6).

Hazir et al. (2016) study the effect of R&D collaboration networks on regional knowledge creation relying on an SDM specification. Their collaboration network is constructed from a convex combination of three connectivity matrices:  $W_c = \gamma_1 W_1 + \gamma_2 W_2 + \gamma_3 W_3$ .

They rely on estimated log-likelihood values to determine an estimate of the true values of  $\gamma_\ell$ ,  $\ell = 1, 2, 3$ , which we designate as  $(\bar{\gamma}_\ell)$ , which is then used to form  $\bar{W}_c = \sum_{\ell=1}^3 \bar{\gamma}_\ell W_\ell$ . Their SDM specification using  $\bar{W}_c$  is then estimated by maximum likelihood.

In this paper, we consider two Bayesian procedures that can be used to estimate the parameters  $\gamma_\ell$ . The first, labeled the *grid approach*, is presented in Section 2.1, which is similar to the approach used in Hazir et al. (2016). We show that it can suffer from a drawback associated with ignoring uncertainty regarding the underlying distribution of  $\gamma_\ell$ . Two possible consequences of this are biases in the scalar summary measures of impacts proposed by LeSage and Pace (2009) and incorrect coverage intervals.

Our second procedure, labeled the *MCMC approach* is presented in Section 2.2. This procedure accounts for uncertainty regarding an underlying distribution for the parameters  $\gamma_\ell$ . In our Monte Carlo experiments, we find that this approach results in centered estimators of the parameters and scalar summary measures as well as superior empirical coverage intervals especially in the case of the scalar summary measures for indirect (spillover) impacts.

## 2.1 The grid approach

Consider the following SAR model where  $\rho_0$ ,  $\beta_0$ ,  $\sigma_0$  and  $\gamma_{\ell,0}$ ,  $\ell = 1, \dots, L$  denotes the true values of the parameters:<sup>4</sup>

$$\begin{aligned}
 y &= \rho_0 W_c y + X \beta_0 + \varepsilon \quad \varepsilon \sim N(0, \sigma_0^2 I_n) \\
 W_c &= \sum_{\ell=1}^L \gamma_{\ell,0} W_\ell \\
 \text{s.t.} \quad & 0 \leq \gamma_{\ell,0} \leq 1, \ell = 1, \dots, L \text{ and } \sum_{\ell=1}^L \gamma_{\ell,0} = 1
 \end{aligned} \tag{7}$$

Let  $\gamma_0 = (\gamma_{1,0}, \dots, \gamma_{L,0})'$ , be the vector containing all parameters of the convex combination. The first approach relies on Bayesian posterior model probabilities that can be calculated for a discrete set of  $M$  models based on a grid of values for each  $\gamma_\ell$ . Each model  $m = 1, \dots, M$  is *conditional* on a particular value for each  $\gamma_\ell$ , denoted  $\gamma_\ell^m$ . For each  $W_c(\gamma^m) = \sum_{\ell=1}^L \gamma_\ell^m W_\ell$ ,  $m = 1, \dots, M$ , we can calculate the log-marginal likelihood of model (7), which is used to compute posterior model probabilities, defined as:  $Prob(M_m | W_c(\gamma) = W_c(\gamma^m))$ . The notation  $Prob(M_m | W_c(\gamma) = W_c(\gamma^m))$  indicates that the posterior model probabilities are conditional on a particular convex combination matrix  $W_c$ , itself determined by the values of the parameter vector  $\gamma^m = [\gamma_1^m, \dots, \gamma_L^m]$ , but not on the model parameters (e.g.,  $\rho, \beta, \sigma^2$ ). The process of calculating log-marginal likelihoods integrates out the model parameters, so the posterior model probabilities are valid for any parameter values of  $\rho, \beta, \sigma^2$  (in the appropriate parameter space) (see LeSage and Pace, 2009, Chapter 6).

An estimator of the true value  $\gamma_0$  that we designate as  $\bar{\gamma}$ , can be determined using the highest posterior model probability from the set of probabilities (each conditional on a particular convex combination of matrices).<sup>5</sup> Formally,  $W_c(\bar{\gamma}) = \sum_{\ell=1}^L \bar{\gamma}_\ell W_\ell$ , where  $\bar{\gamma} = \max_{m=1}^M Prob(M_m | W_c(\gamma) = W_c(\gamma^m))$ . In the sequel, we refer to this approach as the *grid approach* to modeling convex combinations of different types of cross-sectional dependence.

The grid approach should in many cases produce a result equivalent to that of Hazir et al. (2016) who used the maximum of the (conditional) likelihood function across a grid of

<sup>4</sup>We use the SAR model to illustrate issues here, with other cross-sectional dependence model specifications discussed in Appendix A.1.

<sup>5</sup>In this section, we rely on the notation  $\bar{\gamma}$  rather than on  $\hat{\gamma}$  to designate an estimator of  $\gamma$  since, as discussed later, the grid approach ignores the uncertainty embedded in the estimator.



different values for  $\gamma_\ell, \ell = 1, \dots, L$  to determine  $\bar{\gamma}$ . Formally, in the Hazir et al. (2016) paper,  $\bar{\gamma} = \max_{m=1}^M \text{Log}L(M_m | W_c(\gamma) = W_c(\gamma^m))$ , where  $\text{Log}L(M_m | W_c(\gamma) = W_c(\gamma^m))$  denotes the log likelihood function value associated with  $\gamma^m$ .

A possible advantage of the Bayesian grid approach set forth above over that of Hazir et al. (2016) is that the posterior model probabilities are unconditional on the estimated values of the model parameters (e.g.,  $\rho, \beta, \sigma^2$ ). They however remain conditional on the particular  $W_c(\gamma^m)$  matrix used (and therefore conditional on specific values of the convex combination in the grid of values used). Note that in the approach of Hazir et al. (2016), estimates of the SAR parameters  $\rho^m, \beta^m, \sigma^{2,m}$  used to evaluate the likelihood for values of  $\gamma^m, m = 1, \dots, M$  will equal true values for only one of the  $m$  values from the set of  $M$  different values considered. This assumes that we view the correct model as one based on a single value of  $\gamma_0$ , then estimated parameters  $\hat{\rho}, \hat{\beta}, \hat{\sigma}^2$  from models that condition on values of  $\gamma^m \neq \gamma_0$  might lead to incorrect likelihood function values used to determine the estimate  $\bar{\gamma}$ .

It is important to note that we are unable to carry out a formal decomposition of the partial derivative impacts arising from the different types of cross-sectional dependence being modeled. Consider the partial derivative with respect to the  $r^{\text{th}}$  regressor:

$$\partial E(y) / \partial X'_r = (I_n - \hat{\rho} W_c(\bar{\gamma}))^{-1} I_n \hat{\beta}_r \quad (8)$$

$$W_c(\bar{\gamma}) = \sum_{\ell=1}^L \bar{\gamma}_\ell W_\ell \quad (9)$$

$$\partial E(y) / \partial X'_r = \left( I_n - \hat{\rho} \left( \sum_{\ell=1}^L \bar{\gamma}_\ell W_\ell \right) \right)^{-1} I_n \hat{\beta}_r \quad (10)$$

The inverse from the partial derivative expression will involve cross-products of the various  $W_\ell$  matrices, powers of these, as well as the underlying parameter estimates:  $\hat{\rho}, \bar{\gamma}, \hat{\beta}$ . However, we do have estimates of  $\gamma_{\ell,0}, \ell = 1, \dots, L$  that allow us to draw conclusions about the relative importance of each type of connectivity embodied in each of the matrices  $W_\ell$ . To illustrate, consider a convex combination of two matrices:  $W_c = \gamma_1 W_1 + \gamma_2 W_2$ . Due to the convex combination constraint (the sum of the parameters must equal one), we can rewrite this convex combination as  $W_c = \gamma_1 W_1 + (1 - \gamma_1) W_2$ . A value  $\bar{\gamma}_1 = 0.5$  would imply that both types of connectivity are of equal importance, whereas an estimate  $\bar{\gamma}_1 = 0.9$  tells us

that the weight matrix  $W_1$  is much more important than  $W_2$ .

Because the model partial derivatives in (8) reflect an  $n \times n$  matrix, LeSage and Pace (2009) propose a scalar summary measure of the direct effects based on an average of the main diagonal elements of the  $n \times n$  matrix in (8), and a scalar summary of the cumulative indirect effects based on an average of the sum of the off-diagonal elements from each row  $i = 1, \dots, n$  of the matrix in (8). These scalar summaries represent an average of the own-partial derivatives (direct effects =  $\partial y_i / \partial x_i^r$ ) and an average of the cumulative sum of cross-partial derivatives (indirect effects =  $\partial y_j / \partial x_i^r$ ).

Some situations that might arise when using the grid approach are discussed here. For the sake of clarity, we consider again the convex combination  $W_c = \gamma_1 W_1 + (1 - \gamma_1) W_2$ . Suppose that using a 0.1 grid over  $[0, 1]$  for  $\gamma_1$ , we find posterior model probabilities that point to:  $\text{Prob}(M_m | \gamma_1^m = 0.3) = 0.49$  and  $\text{Prob}(M_o | \gamma_1^o = 0.4) = 0.51$ .<sup>6</sup> The question that arises is: should we produce estimates and inferences based on the model with posterior probability equal to 0.51 and ignore estimates and inferences arising from the model based on posterior probability equal to 0.49? A solution to this is to define a finer grid, but the question can still arise, if  $\text{Prob}(M_m | \gamma_1^m = 0.42) = 0.49$  and  $\text{Prob}(M_o | \gamma_1^o = 0.43) = 0.51$ , which model should be used to produce estimates and inferences? This type of result is not likely to be a problem in applied practice because model probabilities like these simply indicate that estimates and inferences from models based on either  $\text{Prob}(M_m | \gamma_1^m = 0.42) = 0.49$  or  $\text{Prob}(M_o | \gamma_1^o = 0.43) = 0.51$  should be very similar.

A related point is that use of a finer grid of values over  $[0, 1]$  for  $\gamma_1$  will likely reduce posterior model probabilities for individual  $\gamma_1^m$  values. This would occur in situations where support for more than a single  $\gamma_1^m$  arises, since posterior model probabilities are spread over the grid of values. One might see a situation where the highest two model probabilities take the form:  $\text{Prob}(M_m | \gamma_1^m = 0.42) = 0.12$  and  $\text{Prob}(M_o | \gamma_1^o = 0.43) = 0.13$ . The decision rule for selecting  $\gamma_1$  would still be selection of  $\bar{\gamma}_1 = 0.43$ , since this is the highest model probability.

Another drawback to selecting a single value of  $\bar{\gamma}$  is that uncertainty regarding the parameter  $\gamma$  is ignored. Indeed, we first select a single value for the parameter vector  $\gamma$  which maximizes the posterior probability and then consider this value as the “true” one when subsequently estimating the spatial regression specification (model (7) here).

<sup>6</sup>We use the notation:  $\text{Prob}(M_m | \gamma_1^m = 0.4)$  as shorthand for:  $\text{Prob}(M_m | W_c(\gamma_1) = W_c(\gamma_1^m), \gamma_1^m = 0.4)$ .

Consequently, we ignore uncertainty regarding this set of parameters. A further point is that uncertainty regarding  $\gamma$  can have consequences due to likely correlation between this parameter vector and other model parameters.<sup>7</sup> When the posterior probability associated with the preferred value, namely  $\bar{\gamma}$ , is very high, the level of uncertainty regarding the value of the parameters  $\gamma$  is very low and ignoring it should have a minor impact on the overall quality of inference. However the situation is different if posterior probabilities associated with other values of  $\gamma_\ell \neq \bar{\gamma}$  are not close to zero. In this case, we face a high level of uncertainty with respect to  $\bar{\gamma}$ .

Since the partial derivatives are non-linear functions of the underlying SAR parameter estimates  $\bar{\gamma}, \hat{\rho}, \hat{\beta}$  as shown in (10), ignoring uncertainty in  $\gamma$  and fixing it at a single estimated value  $\bar{\gamma}$  might produce *poor* coverage intervals for the scalar summary direct and indirect effects estimates. By coverage intervals we mean the interval (say  $\alpha/2\%$  and  $(1 - \alpha/2)\%$  quantiles) determined by our estimates that should encompass the true parameter values  $(1 - \alpha)\%$  of the time. Inference in cross-sectional dependence regression models centers on the direct and indirect effects estimates, for which we require not only point estimates but also measures of dispersion on which inferences about statistical significance of these effects are based.

In section 4 we describe Monte Carlo experiments that demonstrate how taking the grid approach to estimating a convex combination cross-sectional dependence model can in fact produce coverage intervals that are too small. An example of an interval that is too small would be when the estimated 2.5% and 97.5% intervals contain the true parameter only 80% of the time, rather than the expected 95% of the time.

## 2.2 The Bayesian MCMC approach

If there is uncertainty regarding the parameter vector  $\gamma$  on which the grid approach conditions its estimates, the Bayesian solution to this problem is to integrate out these parameters when producing estimates and inferences. Integrating out a model parameter makes posterior inferences *unconditional* on this parameter. By this, we mean that estimates and inferences are valid for all values taken by the parameter over the range of its parameter space.

Our model probabilities are conditional on the (estimated) weight matrix  $\bar{W}_c$ , which, as

---

<sup>7</sup>The variance-covariance matrix of the model parameters thus has a conditional interpretation.

already noted, depends on exogenous fixed information contained in the matrices  $W_\ell$ ,  $\ell = 1, \dots, L$  as well as signal-to-noise in the sample data that gives rise to uncertainty about the parameter vector  $\gamma$ . We could attempt numerical integration of these parameters over the feasible interval ( $0 \leq \gamma_\ell \leq 1$ ),  $\ell = 1, \dots, L$ . However, this would require use of multivariate numerical integration procedures to integrate all model parameters  $\rho, \gamma, \beta, \sigma^2$ . For the case where we condition on a single convex combination matrix  $W_c$  (defined by the values of  $\gamma_\ell$ ), it is possible to analytically integrate out the parameters  $\sigma^2$  and  $\beta$ , and then rely on simple univariate numerical integration over the parameter  $\rho$  to calculate posterior model probabilities. This approach treats the model as *conditional* on a given matrix  $W_c(\bar{\gamma})$ , which allows analytical integration over the parameters  $\beta, \sigma^2$ , and the ensuing univariate numerical integration problem involving  $\rho$  (see LeSage and Pace, 2009, chapter 6). Full treatment of a model that is *unconditional* on the matrix  $W_c(\gamma)$  does not allow analytical integration over the parameters  $\beta, \sigma^2$ , leading to the need for multivariate numerical integration over  $\beta, \sigma^2, \rho$  as well as the parameters in the vector  $\gamma$ . This would be a very computationally challenging problem to solve.

An alternative that we adopt here is to integrate out the parameters  $\gamma$  over the feasible interval ( $0 \leq \gamma_\ell \leq 1$ ) (with  $\sum_{\ell=1}^L \gamma_\ell = 1$ ) using Bayesian Markov Chain Monte Carlo (MCMC) estimation of the model, which also integrates out all other model parameters ( $\rho, \beta, \sigma^2$ ). We will refer to this method for estimating convex combinations of cross-sectional dependence models as the *MCMC approach*. Since this approach incorporates uncertainty regarding all model parameters including those in the vector  $\gamma$ , we would expect that posterior estimates and inferences based on the MCMC approach should produce correct coverage intervals that conform to the desired size (e.g., 95% intervals).

MCMC estimation consists of sampling draws from the complete sequence of conditional posterior distributions for all model parameters. Implementation of MCMC estimation of the parameters  $\gamma_\ell, \ell = 1, \dots, L$  requires that we sample from the conditional distributions of each parameter  $\gamma_\ell | \gamma_{-\ell}, \beta, \rho, \sigma^2$ , where  $-\ell$  stands for all other parameters except the  $\ell^{\text{th}}$ . This sampling occurs in the broader context of sampling for the other parameters  $\beta, \rho, \sigma^2$  based on their conditional distributions as well.

Bayesian estimation typically relies on prior distributions assigned for the model parameters. However, when applying our MCMC approach to estimating parameters  $\beta, \rho, \sigma^2, \gamma$ , we rely on uninformative priors. This involves assigning normal priors for the parameters

$\beta$  in conjunction with Gamma priors for  $\sigma^2$ . Setting the prior variances for the parameters  $\beta$  to very large values essentially eliminates the influence of any prior information on the posterior model estimates for the parameters  $\beta$ . There are also settings that can be used to make the Gamma priors for  $\sigma^2$  have little influence on the posterior estimates for this parameter. For the parameters  $\rho$  and  $\gamma_\ell$ ,  $\ell = 1, \dots, L$ , we rely on proper uniform priors over the stable parameter spaces for these parameters, which also has little to no impact on the posterior estimates for these parameters.

In the next section we set forth conditional distributions required to produce MCMC estimates for the SAR regression model specification. Conditional distributions needed to carry out MCMC estimation of the SLX, SDM, SDEM specifications are presented in Appendix A.1

### 2.3 Conditional distributions for the convex combination SAR model

Conditional distributions for the model parameters required to implement MCMC estimation of the SAR spatial regression specifications in (7) are set forth here.

We rely on a normal prior for  $\beta$ :

$$\pi(\beta) \sim N(\bar{\beta}, \bar{\Sigma}_\beta) \quad (11)$$

where  $\bar{\beta}$  is a  $k \times 1$  vector of prior means and  $\bar{\Sigma}_\beta$  is a  $k \times k$  prior variance-covariance matrix.<sup>8</sup>

We employ a uniform prior for  $\rho$  since this scalar dependence parameter is constrained to lie in the open interval  $(-1, 1)$ .<sup>9</sup> The constraint is imposed during MCMC estimation using griddy Gibbs sampling that integrates over this interval (see LeSage and Pace, 2009, chapter 6). Similarly, we impose the closed interval  $[0, 1]$  for  $\gamma_\ell$ ,  $\ell = 1, \dots, L$  during MCMC estimation, using griddy Gibbs sampling that integrates over this interval. To impose  $\sum_{\ell=1}^L \gamma_\ell = 1$ , we set  $\gamma_L = (1 - \sum_{\ell=1}^{L-1} \gamma_\ell)$  during MCMC sampling.

For the parameter  $\sigma^2$ , we use an inverse Gamma( $\bar{a}, \bar{b}$ ) distribution shown in (12). We note that as values of  $\bar{a}, \bar{b} \rightarrow 0$ , this prior distribution becomes uninformative, which might

<sup>8</sup>We do not separately consider the intercept vector and associated parameter for notational simplicity, assuming this is part of the matrix  $X$  and associated parameter vector  $\beta$ .

<sup>9</sup>A value of -1 is often used in practice as this ensure the matrix inverse:  $(I_n - \rho W_c)^{-1}$  exists. This has the advantage that we do not have to calculate the minimum eigenvalue of  $W_c$  which changes as a function of the values taken by  $\gamma$ .

be important in applied practice since there would be little basis for assigning prior values for the parameter  $\sigma^2$ .

$$\begin{aligned}\pi(\sigma^2) &= \frac{\bar{b}^{\bar{a}}}{\Gamma(\bar{a})}(\sigma^2)^{-(\bar{a}+1)}\exp(-\bar{b}/\sigma^2) \\ \sigma^2 &> 0, \bar{a}, \bar{b} > 0\end{aligned}\quad (12)$$

As is traditional in the literature, we assume that priors for the parameters  $\beta, \rho, \gamma, \sigma^2$  are independent.

The conditional posterior for  $\beta$  (given  $\rho, \gamma, \sigma^2$ ) takes the form in (13). We note that if we wish to eliminate the influence of any prior information on the posterior estimates, we can impose  $\bar{\Sigma}_\beta = \infty$  by modifying (14) to take the form:  $\tilde{\beta} = (X'X)^{-1}X'(y - \rho W_c(\gamma)y)$ .

$$p(\beta|\rho, \gamma, \sigma^2) \sim N(\tilde{\beta}, \tilde{\Sigma}) \quad (13)$$

$$\tilde{\beta} = (X'X + \sigma^2\bar{\Sigma}_\beta^{-1})^{-1}(X'(y - \rho W_c(\gamma)y) + \sigma^2\bar{\Sigma}_\beta^{-1}\bar{\beta}) \quad (14)$$

$$\tilde{\Sigma} = \sigma^2(X'X + \sigma^2\bar{\Sigma}_\beta^{-1})^{-1}$$

The conditional posterior for  $\sigma^2$  (given  $\beta, \rho, \gamma$ ) takes the form in (15), when we set the prior parameters  $\bar{a} = \bar{b} = 0$ .

$$p(\sigma^2|\beta, \rho, \gamma) \propto (\sigma^2)^{-(\frac{N}{2})}\exp\left(-\frac{1}{2\sigma^2}(y - \rho W_c(\gamma)y - X\beta)'(y - \rho W_c(\gamma)y - X\beta)\right) \quad (15)$$

$$\sim IG(\tilde{a}, \tilde{b})$$

$$\tilde{a} = N/2$$

$$\tilde{b} = (y - \rho W_c(\gamma)y - X\beta)'(y - \rho W_c(\gamma)y - X\beta)/2$$

The (log) conditional posterior for  $\rho$  (given  $\beta, \gamma, \sigma^2$ ) takes the form in (16), where we use  $\ln|I_n - \rho W_c(\gamma)|$  to show that the log-determinant term in this model depends on the vector  $\gamma$ . For example, considering a convex combination of 3 matrices, we need to calculate:  $\ln|I_n - \rho W_c(\gamma)| = \ln|I_n - \rho(\gamma_1 W_1 + \gamma_2 W_2 + \gamma_3 W_3)|$  with  $\gamma_3 = 1 - \gamma_1 - \gamma_2$ .

$$\begin{aligned}
p(\rho|\beta, \gamma, \sigma^2) &\propto -\frac{N}{2}\ln\sigma^2 + \ln|I_n - \rho W_c(\gamma)| \\
&\quad - \frac{1}{2\sigma^2} ([I_n - \rho W_c(\gamma)]y - X\beta)' ([I_n - \rho W_c(\gamma)]y - X\beta)
\end{aligned} \tag{16}$$

This distribution does not reflect a known form as in the case of the conditional distributions for  $\beta, \sigma^2$ . We sample the parameter  $\rho$  from this conditional distribution using a griddy Gibbs sampling approach described in detail in LeSage and Pace (2009, chapter 6). This involves univariate integration of the conditional distribution over a grid of values for the parameter  $\rho$ , with a draw-by-inversion from the profile of the cumulative distribution, with details provided in Appendix A.2.1.

The (log) conditional posterior for  $\gamma_\ell$  (given  $\beta, \rho, \sigma^2, \gamma_{-\ell}$ ) takes a form similar to that for  $\rho$ , which has an unknown distributional form. This also requires use of the griddy Gibbs sampling approach, where we perform univariate integration of the conditional posterior expression over the parameter vector  $\gamma$ , using the approach described in detail in Appendix A.2.2. The sampling of each parameter  $\gamma_\ell$  in  $\gamma$  is done conditional on all the other parameters  $\gamma_{-\ell}$ .

$$\begin{aligned}
p(\gamma_\ell | \rho, \beta, \sigma^2, \gamma_{-\ell}) &\propto -\frac{n}{2}\ln\sigma^2 + \ln|I_n - \rho W_c^\ell(\gamma_\ell|\gamma_{-\ell})| - \frac{1}{2\sigma^2} e'e \\
e &= (I_n - \rho W_c^\ell(\gamma_\ell|\gamma_{-\ell}))y - X\beta \\
W_c^\ell(\gamma_\ell|\gamma_{-\ell}) &= \gamma_\ell W_\ell + \sum_{o \neq \ell} \gamma_o W_o
\end{aligned} \tag{17}$$

Of course, when sampling  $\gamma_\ell|\gamma_{-\ell}$ , we use the most recently sampled values of  $\gamma_{-\ell}$ . After sampling an update for each parameter  $\gamma_\ell, \ell = 1, \dots, L$ , the matrix  $W_c(\gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$  is recomputed using the most recently updated value of the vector  $\gamma$ .

## 2.4 The MCMC sampler

MCMC sampling involves evaluating each of these conditional distributions in sequence a large number of times, say  $T$  with some initial number  $S$  excluded from the sample of MCMC draws, leaving us with  $T - S$  draws for each model parameter. Posterior means and measures of dispersion are based on the  $T - S$  retained MCMC draws.

Sampling begins with arbitrary values for the parameters  $\beta, \rho, \gamma, \sigma^2$  for the SAR model, arbitrary values for  $\beta, \theta, \rho, \gamma, \sigma^2$  in the case of the SDM specification,  $\beta, \theta, \lambda, \gamma, \sigma^2$  for the SDEM, and  $\beta, \theta, \gamma, \sigma^2$  for the SLX.

One thing to note about MCMC estimation of the convex combination spatial model versus the conventional single connectivity matrix model is that we need to evaluate the log-determinant term  $\ln|I_n - \rho W_c(\gamma)|$  that appears in the conditional distribution for  $\rho$  and  $\gamma$  on every pass through the MCMC sampler. This is not the case in the conventional single fixed weight matrix  $W$  where this term can be evaluated over a grid of values for the parameter ( $-1 < \rho < 1$ ) once before carry out MCMC iterations.<sup>10</sup> We avoid the need to evaluate the log-determinant term on every MCMC pass by carrying out these calculations once over a grid of values for  $\rho$  and  $\gamma_\ell, \ell = 1, \dots, L$  prior to MCMC sampling and storing the scalar results in an  $np \times ng \times L$  matrix, where  $np$  is the size of the grid of values used for  $\rho$ ,  $ng$  is the size of the grid of values for  $\gamma$  and  $L$  is the size of the parameter vector  $\gamma$ . During MCMC sampling, we lookup the appropriate log-determinant value given the current values of the parameters  $\rho$  and  $\gamma$ .<sup>11</sup>

We need to sample updated values for each of the parameters conditional on values of all others. One pass through the sampler involves producing draws for values of all parameters, and  $T$  of such passes are carried out, with draws from some initial number of passes  $S$  discarded to allow the sampler to “burn-in”. Posterior means, standard deviations, and other summary statistics for the parameter distributions are analyzed using the sample of  $T - S$  retained draws.

For example, if  $(T - S)$  equals 1000, we use the 1000 MCMC draws to calculate the posterior mean, standard deviation, percentiles, etc. for each model parameter. This reflects an empirical posterior distribution for the parameters. In the case of the non-linear scalar summary direct and indirect effects for the SAR specification, we use the retained 1,000 draws for the parameters  $\gamma, \rho, \beta$  (that we represent as:  $\gamma^d, \rho^d, \beta^d, d = 1, \dots, 1000$ ) to evaluate the  $n \times n$  matrix partial derivative expression in (18) 1000 times. For each of these evaluations, we calculate the scalar summary measures of direct effects (using the average of the main diagonal elements of the  $n \times n$  matrix), and the cumulative scalar summary

<sup>10</sup>Alternatively, if one relies on the computation of eigenvalues of the connectivity matrix to calculate the log-determinant, we only need to compute them once (see Ord, 1975).

<sup>11</sup>A grid based on 0.01 values for  $\rho$  and  $\gamma_\ell, \ell = 1, \dots, L$  can be used resulting in 199 values for  $-1 < \rho < 1$  and 101 values for  $0 \leq \gamma_\ell \leq 1$  in this case. Spline interpolation is used during the MCMC sampling to produce a finer grid of log-determinant values.



measure of indirect effects (using the average of the sum of off-diagonal elements from the  $n \times n$  matrix). Empirical measures of dispersion can be constructed from the standard deviation or percentiles of the 1000 estimates of direct and indirect effects. For the  $d^{\text{th}}$  draw, equation (18) shows the matrix partial derivative for the  $r^{\text{th}}$  regressor:

$$\left[ \frac{\partial E(y)}{\partial X_r'} \right]^d = (I_n - \rho^d(W_c(\gamma^d)))^{-1} I_n \beta_r^d, \quad d = 1, \dots, 1000 \quad (18)$$

$$W_c(\gamma^d) = \sum_{\ell=1}^L \gamma_\ell^d W_\ell$$

### 3 Guidance for use of convex combinations of connectivity structures

There are some points to consider when implementing the convex combination of  $W$  matrices models. We first examine the distinction between connectivity matrices and regressors and interpretative issues that arise when a functional relationship exists between these. We then discuss the impact of correlation between connectivity matrices and tests that have been proposed to assess this. Our last point concerns issues pertaining to inference at the boundary of the parameter space for  $\gamma$ .

#### 3.1 Relation between $W_c$ and the $X$ -variables

One point to consider when using these models is that matrices  $W_\ell$ ,  $\ell = 1, \dots, L$  should not be based on variables that reflect the explanatory variables. The spirit of SAR, SDM, SLX and SDEM regression models is that a (causal) theoretical relationship exists between the dependent variable outcomes (for instance observations on regions or points in space) in the vector  $y$  and the explanatory variables matrix  $X$  describing characteristics of these regions or points in space. Cross-sectional dependence between observations is determined by factors (say  $Z$ ) that are different in nature from the explanatory variables, since they relate to channels of interaction, transmission or more generally connectivity between observations. The relationship between explanatory variables and outcomes are frequently prescribed by economic theory, which allows us to formulate causal relations. Spatial regression models aim to assess the impact of variation in direct (own-region characteristics) on own-region

outcomes, as well as indirect effects arising from changes in own-region characteristics on other-region outcomes. In the absence of cross-sectional dependence, indirect (spillover) effects are zero, by definition, while direct effects are equal for all observations.

If theory suggests a relationship between explanatory variables and the dependent variable, we should not entertain using these variables to construct dependence structures. This is because the connectivity matrix reflects the dependence mechanism by which regional outcomes are correlated, which captures a different aspect of variation in the dependent variable vector  $y$  than that modeled using the explanatory variables matrix  $X$ .

The objective of including a connectivity matrix  $W_c$  (or simply  $W$  for standard models) is to account for the structure of simultaneous cross-sectional dependence. This dimension of the economic phenomenon under study will determine how changes in own-region characteristics impact dependent variable outcomes in other regions, when we assess the effect of a change in elements of  $x_i$  on outcomes for observation  $y_i$  as well as other observations  $y_j, j \neq i$ . An implication is that the information content of the connectivity matrix and explanatory variables are different, with the weight matrix determining how own- and other-region outcomes are impacted by cross-sectional dependence, whereas explanatory variables reflect an underlying relation between own-region characteristics and own-region outcomes.

Turning to specification of cross-sectional dependence matrices  $W_\ell, \ell = 1, \dots, L$ , Corrado and Fingleton (2012) contend that theory should be useful in determining spatial (and other) weight matrices used in our cross-sectional dependence regression models. However, past applied econometric work has ignored spatial and more general types of cross-sectional dependence, treating cross-sectional observations as independent. This has resulted in a lack of theoretical attention paid to the issue of cross-sectional dependence in spatial econometrics. One exception is peer group dependence in the social network literature where the connectivity matrix may arise from social norm effects (local-average models) or social multiplier effects (local-aggregate effects) depending on its construction (see, among others Ghiglino and Soyal, 2010; Patacchini and Zenou, 2012; Liu et al., 2014). A second exception is in the area of international trade where shares of world income or population of individual countries provide a theoretical basis for dependence (Behrens et al., 2012; Koch and LeSage, 2015).

An important requirement for interpretation of partial derivatives in spatial autoregressive models using the approach of LeSage and Pace (2009) is that we avoid situations where

the matrix  $W_c$  depends on explanatory variables, say the  $r^{\text{th}}$ , which we indicate formally as:  $W_c = f(X^r)$ . A partial derivative invokes the notion of a *ceteris paribus* change in the  $r^{\text{th}}$  explanatory variable impacting the outcomes vector  $y$ , reflected through the (fixed) matrix inverse:  $(I_n - \rho W_c)^{-1}$ . If  $W_c = f(X^r)$ , we cannot draw upon the *ceteris paribus* assumption of LeSage and Pace (2009) because the matrix inverse is no longer fixed.

To properly interpret a model where  $W_c = f(X^r)$ , we would need to calculate how changes in  $X_r$  impact the matrix  $W_c$ , and consider the *total derivative* impact arising from a change in  $X_r$ . Lee and Yu (2012) consider this situation and derive the associated matrix of partial derivatives, shown in (19) for the  $i^{\text{th}}$  observation.

$$\begin{aligned} \frac{\partial E(y)}{\partial x_{i,r}} &= (I_n - \rho W_c(X_r))_{.i}^{-1} \beta_1 + A_i \\ A_i &= \frac{\partial [I_n - \rho W_c(X_r)^{-1}]}{\partial x_{i,r}} (X\beta) \end{aligned} \quad (19)$$

In (19), the expression  $A_i$  consists of an  $n \times n$  matrix,  $\partial[(I_n - \rho W_c)^{-1}]/\partial x_{i,r}$  showing how changes in  $x_{i,r}$  impact this matrix inverse, and of the  $n \times 1$  vector  $X\beta$ , making  $A_i$  an  $n \times 1$  vector.

The  $n \times n$  matrix of partial derivatives used to produce direct and indirect effects in the SAR model for all  $n$  observations now takes the form in (20).

$$\frac{\partial E(y)}{\partial X_r'} = (I_n - \rho W_c(X_r))^{-1} \beta_r + [A_1 \ A_2 \ \dots \ A_n] \quad (20)$$

Calculation of the direct effects would be done by averaging over main diagonal elements and indirect effects by averaging the cumulative sum of off-diagonal elements from each row. However, to implement this in practice would require that we calculate the vectors  $A_i, i = 1, \dots, n$ , which requires knowledge of the functional relationship between  $X_r$  and the matrix  $W_c(X_r)$ , which is unlikely to be known in practice.

By way of conclusion, we can avoid difficulties in applied use of convex combinations of weights if we make an effort to avoid situations where variables used for construction of weight matrices (which we label  $Z$ ) are not confounded with explanatory variables in the matrix  $X$ .

### 3.2 Structures of dependence reflected by weight matrices

A related point is that the matrices  $W_\ell$ ,  $\ell = 1, \dots, L$  should reflect different types of dependence structures, that convey independent information. Intuitively, if  $W_1$  and  $W_2$  are very similar, then a model that attempts to model these very similar types of dependence using a convex combination based on an estimate  $\hat{\gamma}$  is not likely to succeed. To see this, consider again the SAR model with  $W_c = \gamma_1 W_1 + (1 - \gamma_1) W_2$ . When  $W_1 = W_2$ , the scalar parameter  $\gamma_1$  based on values of  $\gamma_1 = 0, 0.5, 1$  would all produce the same likelihood function values, indicating a lack of identification for this parameter.

LeSage and Pace (2014) propose scalar summaries of similarity between spatial lags that result from use of two weight matrices based on the correlation between the two  $n \times 1$  vectors,  $W_1 u$  and  $W_2 u$  for normalized  $W_i$ ,  $i = 1, 2$  using a single  $n \times 1$  vector of independent identically distributed standard normal deviates  $u$ . Ideally, we would like to have a low correlation between  $W_1 u$  and  $W_2 u$ . This measure does not compute the correlation between  $W_1$  and  $W_2$  on an element-by-element basis, but rather compares, for each observation, the weighted sum (assuming weight matrices are row-normalized) of elements from the vector  $u$  obtained with the two matrices. When matrices are different, we expect the weighted sums to be relatively uncorrelated. Of course, in the case of more than two weight matrices, we can examine the correlation matrix for the vectors:  $(W_1 u, W_2 u, \dots, W_L u)$ .

### 3.3 Inference at the boundary of the parameter space for $\gamma$

Another issue that arises is that of appropriate inference when one or several elements of the  $\gamma$  vector lie near (or on) the boundary of their parameter space, i.e., 0 or 1. Producing point estimates and measures of dispersion (that are needed for inference) when parameters are close to the boundaries (say 0.95) introduces a host of issues related to estimation and inference at the boundary of the parameter space.<sup>12</sup> To avoid these issues one can use the grid approach in a first step to determine if values of  $\gamma_\ell$ ,  $\ell = 1, \dots, L$  are near the 0 or 1 boundary values. In practice, if the grid approach points to values of  $\gamma_\ell$  near zero, this suggests the associated  $W_\ell$  matrix does not play an important role in the overall connectivity scheme, represented by  $W_c$ . This first step would convey substantial information regarding

---

<sup>12</sup>Essentially, the distribution of the parameters of interest will be truncated at the boundary value, so estimates of dispersion/variance need to be adjusted to account for truncation.

which matrices are relevant channels from which cross-sectional dependence arises.<sup>13</sup> In contrast, if an element of  $\gamma$  is close to 1, say  $\gamma_k$ , the convex combination approach is not very useful since most information regarding cross-sectional dependence is conveyed by  $W_k$ . Once all irrelevant connectivity matrices have been eliminated (i.e those with an associated parameters close to zero), we can use the *MCMC approach* to estimate the remaining  $\gamma$  as well as the model parameters and draw associated inferences.

## 4 Monte Carlo experiments

In this section we compare estimates obtained from the grid and MCMC approaches using traditional performance measures of bias and precision as well as coverage intervals. Special consideration is given to estimator performance in situations where the parameter  $\gamma$  of the convex combination is close to or at the boundary of the  $[0, 1]$  permissible interval for this parameter. For these experiments, we consider two sample sizes:  $n = 200$  and  $n = 300$ .

The DGP we consider is presented in equation (21),

$$\begin{aligned} y &= \rho W_c y + \beta_0 \iota_n + \beta_1 x_1 + \varepsilon \\ y &= (I_n - \rho W_c)^{-1} (\beta_0 \iota_n + \beta_1 x_1 + \varepsilon) \\ W_c &= \gamma_1 W_1 + (1 - \gamma_1) W_2 \end{aligned} \tag{21}$$

with  $y$ , the  $(n \times 1)$  vector of the dependent variable,  $\iota_n$  the unit vector of dimension  $n \times 1$  and  $x_1$  is drawn from a centered Normal distribution with a standard deviation of 1.5. The parameter of the constant term is set to 0.5 while  $\beta_1 = 1$ . The connectivity matrix  $W_c$  is constructed as a convex combination of two matrices,  $W_1$  and  $W_2$ . Two cases are considered regarding the relation between  $W_1$  and  $W_2$ . We first assume independence between them and construct  $W_1$  using a binary five nearest neighbors definition, while  $W_2$  is based on the (binary) eight nearest neighbors. To ensure independence, we use two independent random vectors of latitude-longitude coordinates to construct the matrices. In addition, we examine performance of the two estimation approaches when  $W_1$  and  $W_2$  are correlated. Correlation between the two  $W$ -matrices was produced by using the same set of latitudes and longitudes to construct  $W_1$  and  $W_2$ , with the matrix  $W_1$  based on 5 nearest neighbors and the matrix  $W_2$  constructed using the 4<sup>th</sup> to 8<sup>th</sup> nearest neighbors. This produces two

---

<sup>13</sup>We note however that we cannot perform inference on elements of  $\gamma$  with this approach.

weight matrices that share common neighbors in 40% of the row elements (e.g., the 4<sup>th</sup> and 5<sup>th</sup> nearest neighbors). According to the scalar summaries of similarity between spatial lags described above, the correlation between  $W_1u$  and  $W_2u$  is 0.354, where  $u$  is a  $n \times 1$  vector drawn from a standard normal distribution.

The error term  $\varepsilon$  is assumed Normally distributed, centered around zero and with a variance  $\sigma^2I$  set so that the signal to noise ratio ( $SNR$ ) of the model is kept constant (see Pace et al., 2012). If we let  $A = (I - \rho W_c)^{-1}X\beta$ ,  $X = [\iota_n, x_1]$ ,  $\beta = [\beta_0, \beta_1]'$  and  $B = (I - \rho W_c)^{-1}$ , the SNR is defined as follows:

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

As such, we know that  $SNR$  will lie between 0 and 1. Setting  $SNR$  to a fixed value, we can solve for  $\sigma^2$ :

$$\sigma^2 = \frac{A'A(1 - SNR)}{tr(B'B)(SNR)}$$

Two different values for  $SNR$  were considered, 0.3 and 0.7, reflecting weak and strong fit of the model to the generated sample data. The parameter  $\rho$  takes on three different values: 0.3, 0.5, 0.7, to cover weak, moderate and strong cross-sectional dependence. Finally, we consider the following values for  $\gamma_1$ , our parameter of interest: 0, 0.1, 0.5, 0.9, 1. As such, we cover values in the interior of the parameter space, as well as values close to the boundaries and at the upper and lower boundaries of the  $[0, 1]$  parameter space. In these simulations, the vector  $x_1$  is fixed across the 1000 replications. For each MCMC replication, we produce 2000 draws for each parameter and disregard the 1000 first draws as burn-in. For the grid approach, we compute the log-marginal likelihood and associated probabilities for values of  $\gamma_1 \in [0, 1]$  using an increment of 0.05 (21 different values).

Tables 1 to 4 report the results for  $n = 200$ .<sup>14</sup> Table 1 summarizes the results for independent  $W_1$  and  $W_2$  matrices and a  $SNR$  of 0.3 while Table 2 results are based on a  $SNR$  of 0.7. Table 3 presents the results for a  $SNR$  of 0.3 and correlated connectivity matrices. Finally, Table 4 shows results for correlated  $W_1$  and  $W_2$  with  $SNR = 0.7$ . All Tables contain a left panel that summarizes results based on the grid estimation approach while the right panel shows outcomes for the MCMC estimation approach.

<sup>14</sup>The results for  $n = 300$  are available upon request to the authors. The conclusions for this larger sample size corroborate those obtained for  $n = 200$ , but with improved performance.

For each of the estimation approaches, we consider several statistical measures of the performance for the parameters  $\gamma_1$ ,  $\rho$ ,  $\beta_1$ , as well as the average direct effect of  $x_1$  ( $dir_1$ ), and average indirect effect ( $ind_1$ ). We calculated two measures of bias, one labeled  $Bias_{av}$  represents the difference between the average value over all replications of the estimate and the true value of the parameter. A second bias measure, labeled  $Bias_{me}$ , was calculated using the median difference of actual versus estimates. We also calculated two measures of dispersion, the root mean square error ( $RMSE$ ) and median absolute deviation ( $MAD$ ). Finally, we study the coverage interval ( $cover.$ ) of the estimates, defined as the (2.5% and 97.5% quantiles), which should encompass the true parameter values 95% of the time. Specifically, this involved a binary variable that takes the value 1 if the true parameter is part of the coverage interval and 0 otherwise, for each replication. Reported coverage statistics are an average of the binary variable over all replications which indicates the percentage of outcomes when the true parameter value lies within the 95% coverage interval.

To compute the average direct and indirect effect of  $x_1$ , we rely on expressions (22) to (24).

$$\frac{\partial y}{\partial x_1'} = S_1(W_c) = (I_n - \rho W_c)^{-1} \beta_1 \quad (22)$$

We then compute the *direct* and *indirect* values as follows:

$$dir_1 = \frac{1}{n} \text{tr}[S_1(W_c)] \quad (23)$$

$$ind_1 = \frac{1}{n} \iota_n' [S_1(W_c) - \text{Diag}(S_1(W_c))] \iota_n \quad (24)$$

where  $\text{tr}(\cdot)$  is the trace operator and  $\text{Diag}(A)$  is a diagonal matrix containing the diagonal elements of the matrix  $A$ .

The discussion on the Monte Carlo (MC) results turns first to measures of bias and dispersion for the two estimation procedures, and then presents results pertaining to coverage intervals. The MC results in Table 1 are for the case of low SNR and independent connectivity matrices. According to the two bias statistics used, the estimator of  $\gamma_1$  obtained by the grid approach does not seem to be affected, when the true value is close or at the boundary of its parameter space. Indeed, when  $\gamma_1 = 0$  and  $\rho = 0.5$ ,  $Bias_{av}$  and  $Bias_{me}$  for  $\gamma_1$  are quite small. The same is true when  $\rho = 0.7$ . However, we note that the bias for  $\gamma_1$  (measured by  $Bias_{av}$ ) is higher in the presence of weak cross-sectional dependence while the

statistic  $Bias_{me}$  seems less affected. For example, when  $\gamma_1 = 1$ ,  $Bias_{av} = -0.142$ . Further, we note that when  $\gamma_1$  is on the  $[0, 1]$  boundaries of the parameter space, the estimator of  $\gamma_1$  obtained by the grid approach is less biased than for the MCMC approach. In the latter approach, we rely on a draw for the parameter  $\gamma_1$  using inversion, based on the integrated (log) pdf (probability density function) of the conditional distribution for this parameter. Given the truncation of the conditional distribution at the  $[0, 1]$  boundaries, this approach does not appear to work well to produce an unbiased posterior mean estimate. However, as soon as the value of  $\gamma_1$  departs from the boundaries of its parameter space, for instance,  $\gamma_1 = 0.1, 0.9$ , estimates based on the MCMC approach result in more acceptable bias levels ( $Bias_{av} = -0.047$  for  $\gamma_1 = 0.9$  and  $\rho = 0.7$ ). Despite this, we find more biased estimates than those resulting from the grid estimation procedure. Turning to the estimates of precision, the MCMC approach is more efficient than the grid in most cases, and precision estimates based on MAD statistic are systematically lower than those based on RMSE.

Turning attention to behavior of estimates for the other parameters of the model, we see that for all values of  $\gamma_1$ , the MCMC approach provides less biased and more efficient estimates than the grid approach. This is especially true regarding indirect effects. For instance, in Table 1 when  $\gamma_1 = 0.5$  and  $\rho = 0.7$ , the  $Bias_{av}$  statistic for the grid approach is 22.275 while it is only of -0.301 for the MCMC procedure. This result also holds for the direct effect, with a smaller discrepancy in outcomes. These differences are likely a result of the fact that the MCMC estimation approach integrates out the parameter  $\gamma_1$  while the grid estimation procedure treats this parameter as fixed. The  $Bias_{me}$  statistic appears to produce better results regarding indirect effects, for both approaches, but the MCMC approach still clearly dominates the grid procedure. Dispersion statistics for indirect effects may take on very large values in the grid approach, as indicated by the RMSE statistic. By construction, this result is partly driven by bias of the two estimators, but the MAD statistic suggests a much higher dispersion for the grid approach than for the MCMC procedure. Considering  $\bar{\gamma}_1$  as the true value in the grid approach (and thus ignoring uncertainty regarding this parameter) has consequences for estimation of the direct and indirect effects, which represent the focus of inference in spatial regression models.

Ignoring uncertainty in the estimation of  $\gamma_1$  also has some consequence for accuracy of inferences regarding the parameters. In the Monte Carlo experiments, we use the coverage interval to explore this issue. The last columns in the left and right panels of Table 1



report the proportion of the true values lying in the 95% interval constructed from the posterior distribution of each parameter, which should equal 0.95. Coverage results for the MCMC approach are good, even for the low SNR (0.3) in this Table, and the relatively small sample size. Indeed, except for the extreme cases where  $\gamma_1$  is at the boundaries of the parameter space, we observe for  $\gamma_1, \rho, \beta_1, dir_1$  coverage intervals close to 95%, while coverage is sometimes smaller for indirect effects. The MCMC approach, by accounting for the uncertainty in the estimation of  $\gamma_1$ , provides accurate coverage intervals. In contrast, coverage intervals for the grid approach depart from 95%, especially for the parameter  $\rho$  and the average indirect effects, and to a lesser extent for the average direct effect. However, the coverage interval for  $\beta_1$  seems unaffected and lies near the correct 95% level. For instance, when  $\gamma_1 = 0.9$  and  $\rho = 0.7$ , coverage intervals for  $\rho, \beta_1, dir_1$  and  $ind_1$  equal 53.1%, 93.2%, 86.8% and 56.6% respectively, lower than the expected 95%. We note that since  $\bar{\gamma}_1$  is treated as fixed by the grid estimation approach, we cannot calculate a coverage interval for this parameter.

Table 2 summarizes results when the signal to noise ratio equals 0.7, and for independent  $W_1$  and  $W_2$ . We first observe that when the data contains more information regarding the dependent variable, bias and dispersion decrease. For the grid approach, the  $Bias_{av}$  statistic for  $\gamma_1$  is much lower, especially for weak cross-sectional dependence. For instance, when  $\gamma_1$  is set to 0 and  $\rho = 0.3$ , we observe a  $Bias_{av}$  statistic of 0.099, compared to 0.168 in Table 1. Biases for the model parameters are also lower, despite some high values ( $Bias_{av} = 13.136$  for the indirect effects when  $\gamma_1 = 0.5$  and  $\rho = 0.7$ ). They nevertheless improve relative to the low signal to noise ratio as we would expect. The picture is the same for the MCMC approach. Lower biases and better precision for the model parameters. Also, as in Table 1, we observe a bias decrease for  $\gamma_1$  when cross-sectional dependence increases. For instance, considering  $\gamma_1 = 0.1$ , the estimator bias when  $\rho = 0.3, 0.5, 0.7$  is respectively 0.164, 0.052 and 0.009. Even though the coverage intervals improve for the grid approach, they nevertheless remain far from the correct 95% levels, while coverage results from the MCMC estimation procedure remain close to the correct levels.

Tables 3 and 4 present Monte Carlo results for the case where  $W_1$  and  $W_2$  are correlated, with Table 3 summarizing outcomes for a low signal to noise ratio and Table 4 reflecting the signal to noise ratio set to 0.7. In Table 3, we observe larger biases (no matter the statistic used) for  $\gamma_1$ , as would be expected. Indeed, as the information content of the two

weight matrices overlap, uncertainty regarding the true value of  $\gamma_1$  increases, leading to less accurate estimates of this parameter by both methods. Again, the MCMC approach results in lower bias for all model parameters than the grid approach (based on both  $Bias_{av}$  or  $Bias_{me}$ ), and calculated measures of dispersion are generally smaller. Coverage intervals are also adversely impacted by the correlation between the two weight matrices, with the impact less for the MCMC approach (the coverage intervals for  $\rho$  and indirect effects varies between 83 and 95 percent).

Table 4 presents results from an experiment where the signal to noise ratio equals 0.7. By comparison with results from Table 3, we see smaller biased for all parameter estimates and less dispersion around the true values. Coverage intervals are higher for the grid approach, but, with the exception of  $\beta_1$ , these remain far from the correct 95% level. For the MCMC approach, we observe a degradation in accuracy of the coverage interval for indirect effects, which ranges between 86 and 96 percent. Comparing these results to those in Table 2, we see that correlation between  $W_1$  and  $W_2$  diminishes performance of the estimates for  $\gamma_1$ , both in terms of bias and precision. In contrast, estimation performance of the other model parameters is similar to that for the case of no correlation between the weight matrices.

Table 1:  $n = 200$ , Signal to-noise ratio = 0.3, independent  $W_1$  and  $W_2$

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\gamma_1$	0.000	0.168	0.050	0.301	0.050	-	0.304	0.293	0.330	0.103	0.000
$\rho$	0.300	0.102	0.149	0.274	0.120	0.724	-0.051	-0.043	0.160	0.109	0.939
$\beta_1$	1.000	-0.008	-0.010	0.120	0.082	0.947	0.006	0.001	0.118	0.083	0.951
$dir_1$	1.012	0.021	0.021	0.125	0.085	0.951	0.006	0.002	0.119	0.084	0.952
$ind_1$	0.417	1.214	0.573	2.762	0.519	0.735	0.028	-0.024	0.320	0.199	0.940
$\gamma_1$	0.000	0.060	0.000	0.122	0.000	-	0.144	0.123	0.162	0.037	0.000
$\rho$	0.500	0.180	0.200	0.227	0.075	0.600	-0.010	0.001	0.114	0.071	0.955
$\beta_1$	1.000	-0.014	-0.017	0.134	0.094	0.936	0.009	0.002	0.126	0.088	0.953
$dir_1$	1.039	0.071	0.069	0.172	0.103	0.935	0.006	-0.000	0.130	0.091	0.962
$ind_1$	0.961	5.036	2.517	9.402	1.824	0.633	0.212	0.118	0.586	0.300	0.961
$\gamma_1$	0.000	0.030	0.000	0.060	0.000	-	0.078	0.068	0.086	0.019	0.000
$\rho$	0.700	0.167	0.180	0.177	0.035	0.370	-0.002	0.005	0.074	0.050	0.971
$\beta_1$	1.000	-0.019	-0.020	0.146	0.096	0.945	0.012	0.005	0.143	0.100	0.950
$dir_1$	1.105	0.258	0.233	0.360	0.160	0.905	0.003	-0.001	0.156	0.110	0.962
$ind_1$	2.228	25.284	17.725	36.333	10.776	0.435	0.712	0.439	1.497	0.688	0.975
$\gamma_1$	0.100	0.125	0.050	0.298	0.150	-	0.240	0.241	0.270	0.099	0.978
$\rho$	0.300	0.073	0.126	0.273	0.129	0.762	-0.068	-0.066	0.163	0.105	0.935
$\beta_1$	1.000	-0.004	0.000	0.120	0.084	0.952	0.006	0.001	0.117	0.082	0.953
$dir_1$	1.009	0.022	0.026	0.125	0.086	0.954	0.006	0.002	0.118	0.083	0.953
$ind_1$	0.419	1.169	0.482	2.950	0.507	0.769	-0.011	-0.065	0.307	0.186	0.943
$\gamma_1$	0.100	0.008	-0.050	0.129	0.050	-	0.095	0.071	0.134	0.055	0.988
$\rho$	0.500	0.152	0.168	0.210	0.092	0.708	-0.050	-0.041	0.132	0.077	0.957
$\beta_1$	1.000	-0.013	-0.014	0.127	0.084	0.949	0.007	0.002	0.126	0.087	0.951
$dir_1$	1.031	0.062	0.054	0.161	0.098	0.950	0.004	-0.001	0.129	0.091	0.954
$ind_1$	0.969	5.547	1.937	11.069	1.587	0.735	0.052	-0.046	0.530	0.292	0.959
$\gamma_1$	0.100	-0.021	-0.050	0.085	0.050	-	0.025	0.009	0.063	0.032	0.994
$\rho$	0.700	0.139	0.153	0.158	0.047	0.626	-0.054	-0.048	0.105	0.058	0.961

Continued on next page

Table 1 – continued from previous page

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\beta_1$	1.000	-0.020	-0.022	0.155	0.108	0.925	0.011	0.006	0.145	0.101	0.951
$dir_1$	1.079	0.180	0.172	0.280	0.145	0.929	0.003	-0.002	0.155	0.109	0.961
$ind_1$	2.255	20.578	13.626	29.148	9.782	0.670	0.176	-0.140	1.328	0.626	0.963
$\gamma_1$	0.500	-0.014	-0.025	0.281	0.175	-	-0.013	-0.018	0.119	0.069	0.986
$\rho$	0.300	0.061	0.087	0.255	0.135	0.827	-0.088	-0.088	0.162	0.091	0.941
$\beta_1$	1.000	-0.002	0.000	0.115	0.076	0.954	0.005	0.002	0.117	0.082	0.954
$dir_1$	1.007	0.021	0.025	0.119	0.075	0.963	0.007	0.004	0.118	0.083	0.953
$ind_1$	0.422	1.303	0.358	3.315	0.475	0.831	-0.060	-0.123	0.289	0.148	0.941
$\gamma_1$	0.500	0.011	0.000	0.204	0.100	-	0.002	-0.001	0.139	0.098	0.965
$\rho$	0.500	0.117	0.136	0.214	0.115	0.800	-0.118	-0.117	0.176	0.085	0.897
$\beta_1$	1.000	-0.010	-0.008	0.127	0.084	0.940	0.007	0.004	0.126	0.089	0.952
$dir_1$	1.021	0.050	0.050	0.157	0.094	0.954	0.005	0.002	0.128	0.091	0.952
$ind_1$	0.979	6.835	2.523	13.122	2.585	0.804	-0.155	-0.293	0.565	0.256	0.905
$\gamma_1$	0.500	0.017	0.000	0.142	0.100	-	0.007	0.004	0.128	0.089	0.940
$\rho$	0.700	0.104	0.135	0.152	0.058	0.803	-0.140	-0.135	0.187	0.087	0.840
$\beta_1$	1.000	-0.012	-0.009	0.147	0.098	0.954	0.012	0.007	0.148	0.103	0.951
$dir_1$	1.049	0.128	0.110	0.237	0.131	0.960	0.003	-0.001	0.155	0.111	0.953
$ind_1$	2.285	22.275	17.227	30.494	11.705	0.838	-0.301	-0.742	1.547	0.609	0.849
$\gamma_1$	0.900	-0.107	0.000	0.281	0.100	-	-0.246	-0.233	0.277	0.095	0.976
$\rho$	0.300	0.116	0.144	0.264	0.106	0.725	-0.030	-0.018	0.142	0.088	0.955
$\beta_1$	1.000	-0.014	-0.011	0.118	0.076	0.947	0.007	0.005	0.117	0.082	0.957
$dir_1$	1.014	0.016	0.020	0.121	0.081	0.957	0.006	0.004	0.119	0.084	0.955
$ind_1$	0.415	1.202	0.458	3.757	0.392	0.731	0.056	0.015	0.308	0.177	0.959
$\gamma_1$	0.900	-0.001	0.050	0.127	0.050	-	-0.107	-0.091	0.133	0.047	0.981
$\rho$	0.500	0.157	0.161	0.205	0.079	0.604	-0.011	-0.008	0.099	0.063	0.973
$\beta_1$	1.000	-0.014	-0.017	0.127	0.087	0.941	0.010	0.006	0.124	0.087	0.957
$dir_1$	1.046	0.070	0.065	0.160	0.095	0.933	0.005	0.002	0.130	0.092	0.957
$ind_1$	0.954	4.001	1.168	8.917	0.869	0.656	0.199	0.074	0.605	0.270	0.972
$\gamma_1$	0.900	0.020	0.050	0.089	0.050	-	-0.047	-0.037	0.072	0.032	0.991
$\rho$	0.700	0.144	0.149	0.161	0.050	0.464	-0.020	-0.021	0.076	0.047	0.979
$\beta_1$	1.000	-0.027	-0.028	0.139	0.098	0.949	0.013	0.008	0.139	0.096	0.958
$dir_1$	1.118	0.222	0.212	0.316	0.146	0.910	0.001	-0.003	0.156	0.109	0.958
$ind_1$	2.215	18.653	6.748	32.115	5.481	0.536	0.621	0.198	1.754	0.638	0.980
$\gamma_1$	1.000	-0.142	0.000	0.264	0.000	-	-0.306	-0.284	0.331	0.086	0.000
$\rho$	0.300	0.142	0.168	0.254	0.092	0.649	-0.005	0.011	0.138	0.085	0.945
$\beta_1$	1.000	-0.004	0.000	0.115	0.077	0.945	0.007	0.003	0.117	0.082	0.956
$dir_1$	1.017	0.031	0.031	0.126	0.079	0.944	0.006	0.004	0.120	0.084	0.956
$ind_1$	0.411	1.133	0.526	3.345	0.379	0.686	0.109	0.074	0.329	0.182	0.948
$\gamma_1$	1.000	-0.056	0.000	0.109	0.000	-	-0.164	-0.150	0.175	0.035	0.000
$\rho$	0.500	0.192	0.199	0.220	0.066	0.432	0.027	0.031	0.092	0.056	0.952
$\beta_1$	1.000	-0.025	-0.019	0.129	0.080	0.937	0.010	0.006	0.124	0.087	0.956
$dir_1$	1.059	0.082	0.073	0.172	0.100	0.931	0.003	0.001	0.131	0.092	0.955
$ind_1$	0.941	3.819	1.565	8.271	0.965	0.478	0.367	0.242	0.687	0.291	0.957
$\gamma_1$	1.000	-0.032	0.000	0.062	0.000	-	-0.103	-0.096	0.110	0.021	0.000
$\rho$	0.700	0.169	0.171	0.176	0.034	0.181	0.031	0.034	0.067	0.039	0.943
$\beta_1$	1.000	-0.043	-0.043	0.139	0.089	0.943	0.014	0.010	0.135	0.094	0.956
$dir_1$	1.164	0.287	0.257	0.375	0.153	0.871	0.001	-0.004	0.159	0.113	0.957
$ind_1$	2.170	16.784	7.084	29.167	4.331	0.275	1.288	0.904	2.103	0.722	0.956

Table 2:  $n = 200$ , Signal to-noise ratio = 0.7, Independent  $W_1$  and  $W_2$ 

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\gamma_1$	0.000	0.099	0.000	0.203	0.000	-	0.219	0.190	0.244	0.068	0.000
$\rho$	0.300	0.067	0.076	0.156	0.075	0.821	-0.009	-0.000	0.112	0.072	0.944
$\beta_1$	1.000	-0.003	-0.001	0.051	0.035	0.950	0.002	0.000	0.051	0.035	0.949
$dir_1$	1.012	0.007	0.007	0.052	0.036	0.952	0.002	0.001	0.051	0.036	0.954
$ind_1$	0.417	0.271	0.215	0.472	0.196	0.821	0.055	0.037	0.235	0.143	0.951
$\gamma_1$	0.000	0.039	0.000	0.078	0.000	-	0.106	0.094	0.117	0.026	0.000
$\rho$	0.500	0.106	0.107	0.137	0.055	0.721	0.011	0.015	0.080	0.051	0.962
$\beta_1$	1.000	-0.010	-0.012	0.056	0.038	0.940	0.004	0.002	0.054	0.037	0.951

Continued on next page

Table 2 – continued from previous page

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$dir_1$	1.039	0.020	0.020	0.063	0.040	0.948	0.002	0.000	0.056	0.039	0.964
$ind_1$	0.961	0.917	0.629	1.642	0.394	0.730	0.187	0.136	0.417	0.225	0.968
$\gamma_1$	0.000	0.026	0.000	0.052	0.000	-	0.063	0.056	0.070	0.015	0.000
$\rho$	0.700	0.110	0.113	0.127	0.041	0.545	0.009	0.014	0.059	0.038	0.971
$\beta_1$	1.000	-0.016	-0.017	0.063	0.040	0.935	0.006	0.004	0.062	0.042	0.947
$dir_1$	1.105	0.082	0.075	0.125	0.059	0.906	0.001	0.000	0.069	0.046	0.967
$ind_1$	2.228	5.899	2.789	10.891	1.681	0.574	0.568	0.410	1.104	0.516	0.975
$\gamma_1$	0.100	0.044	-0.000	0.190	0.100	-	0.164	0.142	0.201	0.085	0.977
$\rho$	0.300	0.056	0.064	0.146	0.085	0.866	-0.029	-0.025	0.116	0.077	0.938
$\beta_1$	1.000	-0.007	-0.005	0.051	0.035	0.957	0.002	0.000	0.050	0.035	0.949
$dir_1$	1.009	0.003	0.004	0.050	0.035	0.965	0.003	0.000	0.051	0.036	0.955
$ind_1$	0.419	0.255	0.177	0.509	0.212	0.868	0.012	-0.012	0.226	0.145	0.942
$\gamma_1$	0.100	-0.004	-0.050	0.102	0.050	-	0.052	0.035	0.087	0.040	0.986
$\rho$	0.500	0.076	0.077	0.131	0.066	0.826	-0.023	-0.020	0.091	0.057	0.966
$\beta_1$	1.000	-0.008	-0.008	0.053	0.035	0.950	0.004	0.002	0.054	0.037	0.950
$dir_1$	1.031	0.015	0.014	0.057	0.037	0.955	0.002	0.001	0.055	0.039	0.963
$ind_1$	0.969	0.913	0.458	2.193	0.420	0.828	0.052	-0.016	0.383	0.212	0.972
$\gamma_1$	0.100	-0.018	-0.000	0.078	0.050	-	0.009	-0.002	0.051	0.028	0.990
$\rho$	0.700	0.084	0.089	0.118	0.060	0.724	-0.037	-0.032	0.080	0.046	0.973
$\beta_1$	1.000	-0.015	-0.018	0.064	0.041	0.951	0.005	0.003	0.063	0.043	0.946
$dir_1$	1.079	0.067	0.059	0.112	0.055	0.944	0.000	-0.001	0.068	0.047	0.964
$ind_1$	2.255	7.097	2.628	13.289	2.389	0.744	0.107	-0.117	1.020	0.474	0.978
$\gamma_1$	0.500	0.030	0.000	0.246	0.150	-	-0.006	-0.009	0.135	0.094	0.978
$\rho$	0.300	0.038	0.042	0.151	0.097	0.896	-0.061	-0.062	0.124	0.068	0.940
$\beta_1$	1.000	-0.005	-0.006	0.053	0.033	0.937	0.003	0.001	0.050	0.035	0.957
$dir_1$	1.007	0.003	0.003	0.053	0.035	0.937	0.003	0.002	0.051	0.036	0.954
$ind_1$	0.422	0.238	0.143	0.527	0.248	0.890	-0.051	-0.091	0.223	0.120	0.940
$\gamma_1$	0.500	0.007	0.000	0.154	0.100	-	0.005	0.007	0.132	0.091	0.951
$\rho$	0.500	0.065	0.074	0.150	0.092	0.850	-0.080	-0.085	0.134	0.070	0.912
$\beta_1$	1.000	-0.004	-0.003	0.055	0.034	0.932	0.004	0.003	0.054	0.038	0.955
$dir_1$	1.021	0.015	0.014	0.059	0.035	0.947	0.003	0.001	0.055	0.039	0.957
$ind_1$	0.979	1.415	0.581	3.115	0.663	0.853	-0.114	-0.219	0.447	0.229	0.911
$\gamma_1$	0.500	0.012	0.000	0.100	0.050	-	0.004	0.001	0.106	0.068	0.932
$\rho$	0.700	0.090	0.106	0.131	0.059	0.803	-0.097	-0.099	0.144	0.070	0.876
$\beta_1$	1.000	-0.007	-0.007	0.066	0.044	0.928	0.007	0.005	0.064	0.044	0.957
$dir_1$	1.049	0.081	0.071	0.134	0.064	0.929	0.001	-0.001	0.067	0.047	0.963
$ind_1$	2.285	13.136	8.185	20.529	6.430	0.803	-0.175	-0.540	1.365	0.556	0.888
$\gamma_1$	0.900	-0.042	0.050	0.178	0.050	-	-0.172	-0.152	0.202	0.068	0.976
$\rho$	0.300	0.070	0.066	0.135	0.068	0.840	0.002	0.006	0.096	0.060	0.955
$\beta_1$	1.000	-0.003	-0.005	0.051	0.034	0.942	0.003	0.002	0.050	0.035	0.960
$dir_1$	1.014	0.008	0.006	0.052	0.034	0.938	0.003	0.002	0.051	0.036	0.956
$ind_1$	0.415	0.255	0.171	0.472	0.166	0.837	0.067	0.043	0.223	0.124	0.958
$\gamma_1$	0.900	-0.000	0.050	0.111	0.050	-	-0.075	-0.063	0.099	0.039	0.980
$\rho$	0.500	0.083	0.079	0.128	0.064	0.755	0.004	0.003	0.075	0.048	0.967
$\beta_1$	1.000	-0.008	-0.009	0.054	0.036	0.940	0.005	0.003	0.053	0.037	0.960
$dir_1$	1.046	0.021	0.023	0.060	0.039	0.953	0.002	0.001	0.056	0.039	0.958
$ind_1$	0.954	0.779	0.409	1.645	0.376	0.759	0.162	0.080	0.429	0.204	0.971
$\gamma_1$	0.900	0.020	0.050	0.083	0.050	-	-0.036	-0.026	0.062	0.031	0.985
$\rho$	0.700	0.086	0.082	0.114	0.056	0.627	-0.007	-0.008	0.062	0.038	0.976
$\beta_1$	1.000	-0.021	-0.020	0.067	0.041	0.917	0.007	0.004	0.060	0.041	0.960
$dir_1$	1.118	0.078	0.072	0.123	0.054	0.907	-0.000	-0.001	0.069	0.048	0.961
$ind_1$	2.215	5.452	1.479	12.742	1.334	0.646	0.530	0.180	1.422	0.493	0.979
$\gamma_1$	1.000	-0.084	0.000	0.159	0.000	-	-0.225	-0.204	0.243	0.054	0.000
$\rho$	0.300	0.084	0.085	0.136	0.063	0.783	0.027	0.032	0.094	0.055	0.939
$\beta_1$	1.000	-0.010	-0.009	0.050	0.034	0.955	0.003	0.002	0.050	0.035	0.959
$dir_1$	1.017	0.003	0.002	0.050	0.034	0.964	0.003	0.002	0.051	0.036	0.954
$ind_1$	0.411	0.270	0.209	0.439	0.159	0.785	0.119	0.096	0.241	0.124	0.939
$\gamma_1$	1.000	-0.042	0.000	0.083	0.000	-	-0.134	-0.124	0.143	0.028	0.000
$\rho$	0.500	0.106	0.102	0.133	0.049	0.624	0.039	0.039	0.077	0.042	0.928
$\beta_1$	1.000	-0.013	-0.016	0.057	0.038	0.932	0.005	0.003	0.053	0.037	0.959
$dir_1$	1.059	0.025	0.022	0.065	0.042	0.942	0.002	0.000	0.057	0.040	0.959
$ind_1$	0.941	0.800	0.519	1.533	0.285	0.633	0.316	0.248	0.505	0.212	0.931
$\gamma_1$	1.000	-0.027	0.000	0.052	0.000	-	-0.090	-0.083	0.095	0.018	0.000
$\rho$	0.700	0.103	0.101	0.115	0.032	0.401	0.040	0.040	0.062	0.029	0.922

Continued on next page

Table 2 – continued from previous page

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\beta_1$	1.000	-0.022	-0.023	0.065	0.042	0.923	0.008	0.005	0.058	0.040	0.961
$dir_1$	1.164	0.088	0.083	0.123	0.052	0.878	0.001	0.000	0.071	0.050	0.968
$ind_1$	2.170	3.412	1.730	8.258	0.788	0.425	1.095	0.818	1.651	0.529	0.915

Table 3:  $n = 200$ , Signal to-noise ratio = 0.3, correlated  $W_1$  and  $W_2$ 

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\gamma_1$	0.000	0.210	0.150	0.330	0.150	-	0.378	0.381	0.393	0.082	0.000
$\rho$	0.300	0.110	0.131	0.197	0.094	0.793	-0.071	-0.065	0.129	0.071	0.909
$\beta_1$	1.000	-0.011	-0.013	0.119	0.082	0.942	0.006	-0.000	0.118	0.084	0.955
$dir_1$	1.012	0.014	0.010	0.123	0.083	0.953	0.004	0.001	0.120	0.085	0.956
$ind_1$	0.417	0.537	0.413	1.013	0.318	0.807	-0.075	-0.092	0.208	0.119	0.910
$\gamma_1$	0.000	0.142	0.100	0.204	0.100	-	0.258	0.244	0.276	0.067	0.000
$\rho$	0.500	0.154	0.170	0.187	0.070	0.651	-0.079	-0.071	0.124	0.064	0.877
$\beta_1$	1.000	-0.020	-0.017	0.128	0.085	0.949	0.007	0.001	0.127	0.090	0.959
$dir_1$	1.039	0.052	0.053	0.153	0.098	0.956	-0.001	-0.004	0.132	0.095	0.956
$ind_1$	0.961	2.287	1.440	3.744	0.962	0.701	-0.159	-0.197	0.358	0.200	0.882
$\gamma_1$	0.000	0.144	0.100	0.190	0.100	-	0.196	0.184	0.211	0.053	0.000
$\rho$	0.700	0.152	0.162	0.163	0.033	0.455	-0.076	-0.066	0.110	0.055	0.832
$\beta_1$	1.000	-0.016	-0.014	0.145	0.097	0.952	0.011	0.005	0.149	0.106	0.961
$dir_1$	1.100	0.241	0.224	0.340	0.144	0.917	-0.008	-0.011	0.163	0.117	0.954
$ind_1$	2.233	16.383	11.138	22.837	6.572	0.535	-0.388	-0.467	0.789	0.444	0.848
$\gamma_1$	0.100	0.142	0.100	0.293	0.200	-	0.303	0.306	0.322	0.077	0.961
$\rho$	0.300	0.109	0.124	0.188	0.096	0.800	-0.071	-0.067	0.129	0.070	0.907
$\beta_1$	1.000	-0.012	-0.017	0.122	0.080	0.939	0.005	0.000	0.118	0.083	0.959
$dir_1$	1.011	0.013	0.010	0.125	0.084	0.948	0.004	-0.000	0.119	0.085	0.958
$ind_1$	0.417	0.488	0.385	0.789	0.314	0.819	-0.076	-0.093	0.208	0.119	0.912
$\gamma_1$	0.100	0.116	0.100	0.214	0.150	-	0.205	0.194	0.233	0.080	0.949
$\rho$	0.500	0.152	0.166	0.187	0.069	0.651	-0.082	-0.074	0.127	0.065	0.873
$\beta_1$	1.000	-0.012	-0.012	0.128	0.086	0.948	0.008	0.004	0.126	0.089	0.960
$dir_1$	1.038	0.062	0.059	0.156	0.099	0.946	0.001	-0.002	0.131	0.094	0.957
$ind_1$	0.962	2.210	1.480	3.463	0.946	0.691	-0.169	-0.210	0.363	0.204	0.877
$\gamma_1$	0.100	0.120	0.100	0.180	0.100	-	0.150	0.140	0.178	0.068	0.924
$\rho$	0.700	0.147	0.159	0.160	0.036	0.465	-0.077	-0.068	0.112	0.055	0.834
$\beta_1$	1.000	-0.029	-0.031	0.152	0.098	0.944	0.011	0.006	0.148	0.104	0.961
$dir_1$	1.099	0.217	0.200	0.322	0.149	0.918	-0.006	-0.010	0.162	0.117	0.953
$ind_1$	2.234	14.988	11.187	21.037	6.680	0.551	-0.397	-0.469	0.793	0.444	0.840
$\gamma_1$	0.500	0.034	0.050	0.300	0.200	-	0.015	0.009	0.115	0.075	0.998
$\rho$	0.300	0.097	0.112	0.182	0.095	0.802	-0.066	-0.061	0.124	0.069	0.912
$\beta_1$	1.000	-0.013	-0.011	0.121	0.081	0.936	0.005	0.002	0.118	0.084	0.959
$dir_1$	1.012	0.013	0.014	0.123	0.082	0.956	0.005	0.002	0.119	0.085	0.955
$ind_1$	0.417	0.452	0.332	0.884	0.301	0.827	-0.068	-0.084	0.205	0.119	0.916
$\gamma_1$	0.500	0.041	0.050	0.197	0.100	-	0.036	0.042	0.140	0.104	0.984
$\rho$	0.500	0.151	0.166	0.183	0.066	0.652	-0.074	-0.067	0.121	0.063	0.881
$\beta_1$	1.000	-0.013	-0.012	0.126	0.083	0.951	0.008	0.005	0.125	0.088	0.960
$dir_1$	1.041	0.068	0.066	0.157	0.094	0.941	0.003	-0.001	0.130	0.093	0.955
$ind_1$	0.959	1.901	1.317	2.877	0.777	0.699	-0.149	-0.189	0.354	0.208	0.881
$\gamma_1$	0.500	0.048	0.050	0.145	0.100	-	0.045	0.048	0.128	0.089	0.970
$\rho$	0.700	0.145	0.157	0.158	0.036	0.440	-0.064	-0.055	0.100	0.050	0.854
$\beta_1$	1.000	-0.027	-0.031	0.148	0.101	0.940	0.011	0.007	0.142	0.099	0.960
$dir_1$	1.110	0.236	0.216	0.331	0.147	0.897	-0.003	-0.006	0.158	0.113	0.957
$ind_1$	2.223	12.864	9.186	18.044	5.654	0.537	-0.313	-0.383	0.750	0.432	0.862
$\gamma_1$	0.900	-0.093	-0.050	0.243	0.150	-	-0.262	-0.260	0.287	0.087	0.988
$\rho$	0.300	0.127	0.139	0.189	0.090	0.733	-0.044	-0.038	0.115	0.069	0.930
$\beta_1$	1.000	-0.007	-0.010	0.116	0.073	0.952	0.007	0.003	0.118	0.082	0.957
$dir_1$	1.016	0.026	0.023	0.122	0.074	0.950	0.006	0.003	0.120	0.086	0.955
$ind_1$	0.413	0.501	0.422	0.805	0.280	0.750	-0.024	-0.041	0.204	0.126	0.931
$\gamma_1$	0.900	-0.044	0.000	0.156	0.100	-	-0.141	-0.123	0.170	0.061	0.985
$\rho$	0.500	0.165	0.173	0.191	0.062	0.528	-0.037	-0.031	0.093	0.054	0.934
$\beta_1$	1.000	-0.020	-0.019	0.128	0.084	0.958	0.010	0.007	0.124	0.086	0.958

Continued on next page

Table 3 – continued from previous page

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$dir_1$	1.054	0.077	0.077	0.167	0.101	0.929	0.004	0.002	0.131	0.093	0.956
$ind_1$	0.946	1.693	1.299	2.390	0.722	0.583	-0.025	-0.061	0.322	0.200	0.934
$\gamma_1$	0.900	-0.032	0.000	0.115	0.100	-	-0.081	-0.067	0.110	0.047	0.977
$\rho$	0.700	0.153	0.162	0.162	0.030	0.267	-0.030	-0.023	0.068	0.039	0.931
$\beta_1$	1.000	-0.030	-0.029	0.138	0.091	0.942	0.014	0.011	0.136	0.092	0.959
$dir_1$	1.148	0.277	0.269	0.361	0.144	0.869	-0.000	-0.003	0.157	0.111	0.956
$ind_1$	2.185	9.708	6.741	13.608	3.708	0.373	-0.037	-0.098	0.652	0.418	0.941
$\gamma_1$	1.000	-0.160	-0.050	0.271	0.050	-	-0.331	-0.322	0.350	0.085	0.000
$\rho$	0.300	0.132	0.145	0.189	0.086	0.717	-0.035	-0.030	0.112	0.069	0.936
$\beta_1$	1.000	-0.012	-0.008	0.119	0.081	0.948	0.008	0.005	0.117	0.081	0.959
$dir_1$	1.017	0.021	0.027	0.126	0.083	0.953	0.006	0.004	0.119	0.085	0.956
$ind_1$	0.411	0.508	0.417	0.788	0.296	0.731	-0.006	-0.025	0.206	0.131	0.936
$\gamma_1$	1.000	-0.090	-0.050	0.149	0.050	-	-0.197	-0.180	0.214	0.049	0.000
$\rho$	0.500	0.176	0.184	0.197	0.057	0.469	-0.022	-0.017	0.084	0.052	0.940
$\beta_1$	1.000	-0.024	-0.026	0.123	0.078	0.949	0.012	0.008	0.124	0.084	0.960
$dir_1$	1.059	0.080	0.074	0.160	0.090	0.950	0.005	0.002	0.131	0.091	0.960
$ind_1$	0.941	1.711	1.359	2.409	0.670	0.529	0.029	-0.008	0.325	0.204	0.945
$\gamma_1$	1.000	-0.072	-0.050	0.113	0.050	-	-0.134	-0.120	0.146	0.033	0.000
$\rho$	0.700	0.160	0.164	0.166	0.028	0.209	-0.016	-0.010	0.059	0.038	0.944
$\beta_1$	1.000	-0.031	-0.029	0.132	0.088	0.952	0.017	0.013	0.135	0.090	0.958
$dir_1$	1.164	0.304	0.283	0.387	0.146	0.865	0.001	-0.002	0.158	0.110	0.957
$ind_1$	2.170	9.657	6.498	13.758	3.213	0.308	0.104	0.035	0.665	0.429	0.952

Table 4:  $n = 200$ , Signal to-noise ratio = 0.7, correlated  $W_1$  and  $W_2$ 

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$\gamma_1$	0.000	0.148	0.100	0.232	0.100	-	0.294	0.282	0.313	0.075	0.000
$\rho$	0.300	0.064	0.069	0.110	0.058	0.875	-0.036	-0.033	0.089	0.055	0.933
$\beta_1$	1.000	-0.006	-0.008	0.051	0.035	0.946	0.002	-0.000	0.051	0.035	0.958
$dir_1$	1.012	0.004	0.001	0.051	0.036	0.956	0.001	-0.000	0.051	0.037	0.952
$ind_1$	0.417	0.195	0.180	0.299	0.140	0.878	-0.034	-0.043	0.156	0.103	0.930
$\gamma_1$	0.000	0.110	0.100	0.163	0.100	-	0.188	0.172	0.204	0.051	0.000
$\rho$	0.500	0.088	0.093	0.119	0.053	0.755	-0.044	-0.041	0.085	0.048	0.912
$\beta_1$	1.000	-0.010	-0.011	0.055	0.039	0.950	0.003	0.001	0.054	0.038	0.963
$dir_1$	1.039	0.018	0.017	0.061	0.042	0.958	-0.001	-0.003	0.057	0.041	0.955
$ind_1$	0.961	0.597	0.525	0.802	0.321	0.768	-0.088	-0.106	0.266	0.167	0.912
$\gamma_1$	0.000	0.106	0.100	0.144	0.100	-	0.149	0.137	0.163	0.042	0.000
$\rho$	0.700	0.104	0.109	0.119	0.036	0.589	-0.050	-0.045	0.081	0.042	0.865
$\beta_1$	1.000	-0.018	-0.022	0.066	0.042	0.945	0.005	0.002	0.064	0.045	0.963
$dir_1$	1.100	0.088	0.086	0.133	0.063	0.915	-0.008	-0.009	0.073	0.052	0.955
$ind_1$	2.233	4.186	2.745	7.362	1.494	0.617	-0.261	-0.310	0.616	0.367	0.878
$\gamma_1$	0.100	0.101	0.050	0.234	0.150	-	0.232	0.221	0.259	0.082	0.967
$\rho$	0.300	0.056	0.061	0.111	0.065	0.869	-0.040	-0.037	0.091	0.054	0.928
$\beta_1$	1.000	-0.005	-0.006	0.051	0.035	0.948	0.002	-0.000	0.051	0.035	0.957
$dir_1$	1.011	0.005	0.004	0.052	0.034	0.949	0.001	-0.000	0.051	0.037	0.952
$ind_1$	0.417	0.179	0.148	0.298	0.151	0.881	-0.042	-0.051	0.157	0.102	0.924
$\gamma_1$	0.100	0.070	0.050	0.154	0.100	-	0.137	0.121	0.168	0.065	0.961
$\rho$	0.500	0.081	0.086	0.112	0.054	0.795	-0.048	-0.044	0.088	0.048	0.896
$\beta_1$	1.000	-0.010	-0.008	0.056	0.039	0.944	0.003	0.001	0.054	0.038	0.958
$dir_1$	1.038	0.017	0.020	0.061	0.041	0.948	-0.001	-0.002	0.057	0.041	0.955
$ind_1$	0.962	0.547	0.490	0.754	0.332	0.804	-0.101	-0.120	0.271	0.165	0.900
$\gamma_1$	0.100	0.082	0.100	0.138	0.100	-	0.105	0.093	0.133	0.056	0.944
$\rho$	0.700	0.097	0.101	0.112	0.039	0.627	-0.052	-0.047	0.083	0.042	0.853
$\beta_1$	1.000	-0.015	-0.015	0.067	0.043	0.927	0.006	0.003	0.063	0.044	0.965
$dir_1$	1.099	0.084	0.080	0.124	0.062	0.909	-0.007	-0.009	0.072	0.051	0.956
$ind_1$	2.234	3.510	2.437	5.172	1.381	0.655	-0.284	-0.340	0.623	0.353	0.867
$\gamma_1$	0.500	0.020	0.000	0.254	0.150	-	0.014	0.010	0.136	0.098	0.991
$\rho$	0.300	0.047	0.048	0.107	0.067	0.868	-0.042	-0.039	0.092	0.055	0.925
$\beta_1$	1.000	-0.003	-0.004	0.052	0.035	0.944	0.003	0.001	0.050	0.036	0.958
$dir_1$	1.012	0.007	0.005	0.053	0.036	0.951	0.003	0.001	0.051	0.036	0.954

Continued on next page

Table 4 – continued from previous page

Param.	True	Grid Approach					MCMC Approach				
		$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.	$Bias_{av}$	$Bias_{me}$	RMSE	MAD	Cover.
$ind_1$	0.417	0.154	0.128	0.277	0.146	0.876	-0.048	-0.054	0.157	0.100	0.923
$\gamma_1$	0.500	0.023	0.050	0.159	0.100	-	0.025	0.030	0.136	0.092	0.975
$\rho$	0.500	0.073	0.080	0.110	0.056	0.797	-0.048	-0.043	0.088	0.049	0.888
$\beta_1$	1.000	-0.010	-0.009	0.055	0.037	0.950	0.004	0.002	0.054	0.038	0.960
$dir_1$	1.041	0.019	0.020	0.059	0.038	0.955	0.000	-0.001	0.056	0.040	0.958
$ind_1$	0.959	0.485	0.429	0.686	0.305	0.799	-0.101	-0.118	0.271	0.164	0.896
$\gamma_1$	0.500	0.031	0.050	0.115	0.050	-	0.028	0.032	0.110	0.072	0.963
$\rho$	0.700	0.093	0.100	0.108	0.035	0.613	-0.043	-0.038	0.075	0.039	0.882
$\beta_1$	1.000	-0.017	-0.016	0.065	0.043	0.934	0.006	0.003	0.061	0.042	0.960
$dir_1$	1.110	0.085	0.080	0.123	0.057	0.898	-0.004	-0.006	0.070	0.050	0.957
$ind_1$	2.223	2.751	2.113	3.986	1.091	0.642	-0.220	-0.279	0.585	0.341	0.887
$\gamma_1$	0.900	-0.070	0.000	0.210	0.100	-	-0.200	-0.185	0.231	0.080	0.975
$\rho$	0.300	0.056	0.062	0.108	0.056	0.852	-0.019	-0.013	0.082	0.052	0.931
$\beta_1$	1.000	-0.005	-0.005	0.050	0.032	0.944	0.004	0.002	0.050	0.035	0.959
$dir_1$	1.016	0.006	0.006	0.051	0.032	0.955	0.003	0.002	0.051	0.036	0.955
$ind_1$	0.413	0.167	0.150	0.274	0.132	0.855	-0.003	-0.007	0.152	0.099	0.941
$\gamma_1$	0.900	-0.022	0.000	0.121	0.100	-	-0.098	-0.083	0.128	0.049	0.981
$\rho$	0.500	0.084	0.087	0.110	0.049	0.712	-0.018	-0.013	0.066	0.041	0.931
$\beta_1$	1.000	-0.010	-0.010	0.055	0.039	0.949	0.006	0.004	0.053	0.037	0.959
$dir_1$	1.054	0.026	0.026	0.065	0.041	0.940	0.002	0.001	0.056	0.039	0.959
$ind_1$	0.946	0.502	0.449	0.720	0.274	0.722	0.001	-0.011	0.237	0.152	0.943
$\gamma_1$	0.900	-0.016	0.000	0.095	0.050	-	-0.058	-0.045	0.087	0.042	0.976
$\rho$	0.700	0.088	0.091	0.100	0.033	0.531	-0.017	-0.013	0.051	0.031	0.936
$\beta_1$	1.000	-0.017	-0.015	0.063	0.043	0.932	0.008	0.006	0.059	0.040	0.960
$dir_1$	1.148	0.089	0.088	0.121	0.056	0.868	-0.001	-0.002	0.070	0.049	0.966
$ind_1$	2.185	1.890	1.588	2.449	0.765	0.566	-0.005	-0.039	0.484	0.307	0.942
$\gamma_1$	1.000	-0.123	0.000	0.210	0.000	-	-0.260	-0.244	0.280	0.070	0.000
$\rho$	0.300	0.065	0.069	0.107	0.057	0.833	-0.010	-0.004	0.079	0.051	0.934
$\beta_1$	1.000	-0.007	-0.007	0.051	0.034	0.944	0.004	0.003	0.050	0.035	0.958
$dir_1$	1.017	0.005	0.007	0.052	0.035	0.951	0.003	0.002	0.051	0.036	0.956
$ind_1$	0.411	0.185	0.169	0.280	0.137	0.832	0.015	0.011	0.153	0.101	0.940
$\gamma_1$	1.000	-0.071	0.000	0.120	0.000	-	-0.153	-0.139	0.166	0.037	0.000
$\rho$	0.500	0.093	0.097	0.113	0.042	0.665	-0.006	-0.001	0.061	0.039	0.942
$\beta_1$	1.000	-0.011	-0.010	0.055	0.034	0.937	0.007	0.005	0.053	0.037	0.958
$dir_1$	1.059	0.027	0.027	0.063	0.038	0.947	0.003	0.001	0.057	0.039	0.963
$ind_1$	0.941	0.530	0.490	0.654	0.248	0.673	0.048	0.036	0.239	0.151	0.948
$\gamma_1$	1.000	-0.057	-0.050	0.089	0.050	-	-0.106	-0.095	0.116	0.026	0.000
$\rho$	0.700	0.093	0.095	0.101	0.025	0.442	-0.004	-0.001	0.044	0.027	0.944
$\beta_1$	1.000	-0.019	-0.018	0.061	0.040	0.929	0.010	0.008	0.059	0.040	0.960
$dir_1$	1.164	0.089	0.083	0.119	0.055	0.880	0.001	-0.001	0.071	0.049	0.966
$ind_1$	2.170	1.816	1.662	2.145	0.599	0.458	0.131	0.098	0.493	0.297	0.956

## 5 An applied illustration using Texas nursing homes

Our application builds on Blankmeyer et al. (2011) model for the (logged) compensation (salary and benefits) of Texas nursing home CEO's and consists of 856 nursing homes in 2002.<sup>15</sup> They argue that (logged) compensation of nursing home CEO's ( $y$ ) is determined by their boards of directors using the following  $X$ -variables: 1) the patient mix achieved by the director which they measure using the percent of total revenue that each facility earned from private pay, medicaid and medicare reimbursements. The argument of Blankmeyer

<sup>15</sup>The sample of 856 nursing homes represents over 80% of facilities licensed by the state in the same year and each facility's senior administrator is treated as the CEO.

et al. (2011) is that private pay and Medicaid patients are less profitable than Medicare patients, implying a negative relationship with compensation for managers in facilities that rely more on private pay and Medicaid payments,<sup>16</sup> 2) the logarithm of the occupancy rate defined as: (total resident days / 365 x number of beds). Blankmeyer et al. (2011) argue that many Texas nursing homes could operate more profitably if they achieved higher occupancy rates. This implies a positive relationship between this variable and managers compensation, indicating that managers should be better compensated in facilities with fewer empty beds. 3) They include two indicator/dummy variables in the set of explanatory variables: i) an indicator for nursing homes that are operated for profit (= 1) or non-profit (= 0), and ii) whether the nursing home is affiliated with a chain of nursing homes (= 1) or is independently owned (= 0). The motivation for these dummies is that for profit versus non-profit and chain versus independent nursing homes provide differential levels of compensation to managers.

Blankmeyer et al. (2011) show dependence between compensation of one nursing home CEO on compensation of CEOs from peer institutions, where peer institutions were defined using institutions with similar expenditures on the nursing function. The motivation for peer effects in CEO's compensation arises from a salary benchmarking argument. Salary benchmarking in this managerial context would evaluate management skills of one CEO using those of others from peer institutions to provide a benchmark. They found evidence of positive peer group dependence, but did not consider spatial dependence in their study.

However, there is an intuitive motivation for considering spatial dependence in addition to the peer group dependence. Owners of nursing homes who are making compensation decisions are more likely to have information regarding compensation of other nursing home managers located nearby. This would be especially true for nursing homes located in major metropolitan areas such as Austin, Dallas, Houston and San Antonio. These owners are likely to interact with other nursing facility owners in their own city.

Our geographically based connectivity matrix  $W_1$  is a binary matrix based on the six nearest distance neighboring nursing homes while our connectivity matrix based on peers,  $W_2$ , is constructed using expenditures on the nursing function to determine the eight nearest peers. These choices were based on a replication of the exploration by Blankmeyer et al.

---

<sup>16</sup>We use *private pay* and *medicare* revenue proportions as two separate explanatory variables, with medicaid excluded to avoid a linear combination that sums of unity. Private pay proportion of revenue should have a negative sign and medicare proportion a positive sign.



(2011) who explored 60 different models based on 1 to 10 peers as well as six different similarity criteria, with the addition of spatial weight matrices consisting of 1 to 8 nearest neighbors.<sup>17</sup> Our findings were that a peer matrix based on expenditures on the nursing function and eight nearest peers along with a spatial weight matrix based on six nearest neighbors had the highest posterior model probability.<sup>18</sup>

We computed the correlation between the two weight matrices using the method developed in Section 3.2. There was a low correlation between  $W_1u$  and  $W_2u$  equal to 0.0221, suggesting these two types of connectivity are capturing different types of dependence.

The model we estimate is presented in (25).

$$\begin{aligned} y &= \rho W_c y + X\beta + \varepsilon \\ W_c &= \gamma_1 W_1 + (1 - \gamma_1) W_2 \end{aligned} \tag{25}$$

Posterior distributions for parameters of the model were constructed from 6,000 MCMC draws with 10,000 draws carried out, but the first 4,000 omitted for burn-in. Two base models were estimated, one based on the spatial neighbors matrix  $W_1$  alone and a second based on the peer institution matrix  $W_2$  alone. We then estimate the convex combination model relying on both the grid and MCMC approaches developed here. The grid approach estimates of the SAR model relied on a 0.01 grid of values for  $\gamma$ , while the MCMC approach was used to integrate out this parameter from the posterior distribution of model parameters.

## 5.1 Benchmark model results based on $W_1$ and $W_2$ alone

Estimation results for the posterior means, medians along with lower and upper 95% credible intervals for models based on a spatial weight matrix alone ( $\gamma_1 = 1$ ) and a peer weight matrix alone ( $\gamma_1 = 0$ ) are shown in Tables 5 and 6.

The signs of the posterior mean estimates of  $\beta$  as well as the direct and indirect effects for the indicator variable for facilities that are part of a *Chain* are positive, indicating that

---

<sup>17</sup>The similarity criteria considered in Blankmeyer et al. (2011) involved: 1) nursing expenditures, 2) nursing staff, 3) square foot area of the facilities, 4) a combination of nursing expenditures and square foot area, 5) a combination of nursing expenditures and nursing staff, and 6) a combination of nursing staff and square foot area. The bivariate measures of similarity were based on Mahalanobis distance to control for covariation between the two measures.

<sup>18</sup>We note that Blankmeyer et al. (2011) use an SDM model and slightly different explanatory variables than those used here.

Table 5: Estimation results based on  $\gamma_1 = 1$  (spatial dependence)

Variables	lower 2.5%	mean	upper 97.5%	sign/signif
Estimates				
Constant	8.9263	9.8792	10.8517	+
For Profit	0.0208	0.0752	0.1282	+
Chain	0.0221	0.0702	0.1186	+
Share Private pay	-0.1556	0.0269	0.2212	
Share Medicare	0.1553	0.2819	0.4054	+
Log(occupancy rate)	0.6044	0.6418	0.6785	+
$W_1y$	0.0226	0.1114	0.1971	+
Direct effects				
For Profit	0.0208	0.0754	0.1282	+
Chain	0.0221	0.0704	0.1188	+
Share Private pay	-0.1561	0.0271	0.2123	
Share Medicare	0.1558	0.2825	0.4061	+
Log(occupancy rate)	0.6055	0.6431	0.6798	+
Indirect effects				
For Profit	0.0009	0.0095	0.0227	+
Chain	0.0009	0.0090	0.0213	+
Share Private pay	-0.0224	0.0034	0.0324	
Share Medicare	0.0057	0.0356	0.0754	+
Log(occupancy rate)	0.0148	0.0811	0.1548	+

Note: + indicates  $> 0$  at the 5% level.  $W_1$  is the geographically-based connectivity matrix.

compensation is higher for managers of these facilities than independent facilities. The *For Profit* indicator variable also has mean estimates ( $\beta$ , direct, indirect) with positive signs, suggesting that these facilities pay higher compensation to managers. The *Medicare* variable also has positive posterior mean  $\beta$ , direct and indirect effects, suggesting that facilities that derive a higher proportion of revenue from Medicare patients reward managers with higher compensation than nursing home relying more on medicaid patients. The *Private pay* variable coefficient is not significantly different from zero, implying that the effect of this variable is not different from the medicaid share reference. *Occupancy rate* has positive posterior mean  $\beta$ , direct and indirect effects, indicating that managers who achieve higher occupancy rates are rewarded with higher compensation.

Despite the similarity in signs and significance of direct and indirect effects of the explanatory variables, there are substantial differences in the magnitude of direct and indirect effects estimates from the two models. Direct effects are larger for the model based on only a spatial weight matrix, while indirect effects are more than 3 times larger for the model based on peer group weights. These differences stem from the much larger estimate for the dependence parameter  $\rho$  in the peer group model (0.32) compared to the spatial model

Table 6: Estimation results based on  $\gamma_1 = 0$  (peer group dependence)

Variables	lower 2.5%	mean	upper 97.5%	sign/signif
Estimates				
Constant	6.7345	7.4724	8.2280	+
For Profit	0.0332	0.0825	0.1328	+
Chain	0.0115	0.0567	0.1020	+
Share Private pay	-0.1237	0.0487	0.2195	
Share Medicare	0.0885	0.2032	0.3193	+
Log(occupancy rate)	0.4127	0.4640	0.4127	+
$W_2y$	0.2561	0.3247	0.3901	+
Direct effects				
For Profit	0.0340	0.0847	0.1363	+
Chain	0.0118	0.0582	0.1048	+
Share Private pay	-0.1270	0.0500	0.2255	
Share Medicare	0.0908	0.2085	0.3285	+
Log(occupancy rate)	0.4228	0.4762	0.5304	+
Indirect effects				
For Profit	0.0142	0.0378	0.0655	+
Chain	0.0050	0.0260	0.0495	+
Share Private pay	-0.0561	0.0224	0.1042	
Share Medicare	0.0383	0.0931	0.1579	+
Log(occupancy rate)	0.1489	0.2126	0.2856	+

Note: + indicates  $> 0$  at the 5% level.  $W_2$  is the peer-group based connectivity matrix.

(0.11).

## 5.2 Estimation results based on the grid approach

Table 7 presents the posterior model probabilities for each value of  $\gamma_1$  for a 0.01 grid.<sup>19</sup> The highest model posterior probability is reached for  $\gamma_1 = 0.20$ . Model probabilities are monotonically rising up to the value of 0.20 and then falling monotonically. We would expect the highest posterior probabilities should be distributed around a central value, which appears to be the case here.

LeSage (2014) notes that numerical integration of the log-marginal likelihood used to construct the posterior model probabilities requires that a great deal of precision be maintained. Publicly available MATLAB<sup>®</sup> functions to calculate log-marginal likelihoods for both cross-sectional and static panel data spatial regression models including SDM, SDEM, SLX are described in LeSage (2015). The functions provide an option to rely on calculation of an exact log-determinant term or a faster approach that uses a Monte Carlo estimate to approximate the log-determinant (see Barry and Pace, 1999). The log-determinant term

<sup>19</sup>We do not report results beyond  $\gamma_1 = 0.51$  as the posterior probabilities are equal to 0.

arises in the expression involving numerical integration needed to produce the log-marginal likelihood. Results in the third column of Table 7 were based on the slower but more accurate exact log-determinant calculation described in Pace and Barry (1997) rather than the faster Monte Carlo estimate to approximate the log-determinant. The faster Monte Carlo estimate of the log-determinant produced the same result pointing to  $\gamma_1 = 0.20$  as the highest probability model.<sup>20</sup>

Table 7: Posterior model probabilities for a grid of  $\gamma_1$

$\gamma_1$	Probability	Log-marginal likelihood	$\gamma_1$	Probability	Log-marginal likelihood
0	0.00736	-396.46966	0.26	0.03658	-394.86671
0.01	0.00848	-396.32813	0.27	0.03337	-394.95853
0.02	0.00974	-396.18963	0.28	0.02990	-395.06837
0.03	0.01116	-396.05435	0.29	0.02629	-395.19710
0.04	0.01273	-395.92254	0.30	0.02266	-395.34549
0.05	0.01446	-395.79470	0.31	0.01914	-395.51432
0.06	0.01636	-395.67117	0.32	0.01583	-395.70439
0.07	0.01843	-395.55224	0.33	0.01280	-395.91648
0.08	0.02065	-395.43836	0.34	0.01013	-396.15121
0.09	0.02302	-395.33002	0.35	0.00782	-396.40936
0.10	0.02550	-395.22767	0.36	0.00590	-396.69157
0.11	0.02807	-395.13177	0.37	0.00434	-396.99841
0.12	0.03068	-395.04283	0.38	0.00311	-397.33046
0.13	0.03328	-394.96137	0.39	0.00218	-397.68807
0.14	0.03582	-394.88788	0.40	0.00148	-398.07173
0.15	0.03822	-394.82302	0.41	0.00098	-398.48172
0.16	0.04040	-394.76737	0.42	0.00064	-398.91827
0.17	0.04230	-394.72157	0.43	0.00040	-399.38142
0.18	0.04382	-394.68626	0.44	0.00025	-399.87122
0.19	0.04489	-394.66214	0.45	0.00015	-400.38753
<b>0.20</b>	<b>0.04544</b>	<b>-394.64989</b>	0.46	0.00009	-400.93015
0.21	0.04543	-394.65015	0.47	0.00005	-401.49869
0.22	0.04482	-394.66373	0.48	0.00003	-402.09265
0.23	0.04359	-394.69143	0.49	0.00001	-402.71145
0.24	0.04178	-394.73396	0.50	0.00001	-403.35430
0.25	0.03942	-394.79210	0.51	0.00000	-404.02028

Table 8 presents the results for the convex-combination SAR model with  $\gamma_1 = 0.2$ . The posterior means for the parameters  $\beta$  fall between those of the two base models reported in Table 5 and Table 6, as we would expect. We observe  $\beta$  estimates closer to those from the model based on peers only where  $\gamma_1 = 0$ , than a value of  $\gamma = 1$  for the spatial dependence model.

Similar estimates for the coefficients  $\beta$  from the two models imply similar direct effects

<sup>20</sup>The difference in speed between the two approaches is quite substantial; Less than one minute for the approximation and several hours for the exact calculation for this application.

estimates, since as pointed out by LeSage and Pace (2009), the difference between the  $\beta$  estimates and direct effects from SAR models is feedback from spatial neighbors. This arises because the main diagonal terms of the partial derivative matrix  $(I_n - \rho W)^{-1} \beta^r = (I_n + \rho W + \rho^2 W^2 + \dots) \beta^r$  contain non-zero elements for the matrices  $W^2, W^3$ , etc. Non-zero elements arise for the matrix  $W^2$  because each nursing facility  $i$  is a neighbor to its neighbor, and diagonal elements of the matrix  $W^2$  reflect non-zero weights assigned to neighbors to neighbors, those for the matrix  $W^3$  reflect neighbors to neighbors to neighbors, (and so on for higher-order neighbors). In the case of our SAR model based on a convex combination of weights, non-zero elements on the diagonal of the matrix  $W_c^2$  reflect the fact that each nursing home facility is a neighbor to both its spatial and peer group neighboring facilities.

In the presence of positive dependence which arises here since  $\rho = 0.40$ , we will have positive feedback effects leading to larger direct effects magnitudes (in absolute value terms) than the magnitude of the coefficients  $\beta$ . We see this is the case when we compare the five respective  $\beta$  posterior mean estimates (0.0795, 0.0562, 0.0486, 0.1900, 0.4635) to the corresponding direct effects (0.0816, 0.0577, 0.0499, 0.1951, 0.4758).

Despite these similarities in estimates of the parameters  $\beta$  and direct effects for models based on  $\gamma_1 = 0.20$  and  $\gamma_1 = 0$ , indirect effects estimates from these two models are quite different in magnitude. Indirect effects for the model based on  $\gamma_1 = 0.20$  are larger in magnitude in all cases, having magnitudes nearly 50 percent larger than indirect effects from the  $\gamma_1 = 0$  model. This indicates that incorporating geographic as well as peer group dependence in the model leads to different posterior mean estimates of spillover impacts on CEO compensation arising from *both* geographic neighbors as well as peer facilities. We also note that the posterior mean and median values for estimated values, direct and indirect effects are similar, suggesting symmetric posterior distributions.

Our approach also allows a test for the presence of other connectivity matrices in the convex combination. For instance, in addition to  $W_1$ , the six nearest neighbors and  $W_2$ , the eight peers based on (logged) expenditures on the nursing function connectivity matrices, we added a third matrix  $W_3$  that reflects eight peers constructed using the (logged) square foot area of the nursing homes. One might expect a high correlation between the two different peer group matrices  $W_2, W_3$ , since these are both based on measures of facility size. However, the correlation between  $W_2 u, W_3 u$  equals 0.0504, suggesting this is not the case. The correlation between  $W_1 u, W_2 u$  equals 0.0221, and that between  $W_1 u, W_3 u$

Table 8: Estimation results based on  $\gamma_1 = 0.20$ 

Variables	lower 2.5%	Mean	Median	upper 97.5%	sign/signif
Estimates					
Constant	5.7590	6.6405	6.6352	7.5497	+
For Profit	0.0302	0.0795	0.0791	0.1296	+
Chain	0.0112	0.0562	0.0563	0.1013	+
Share Private pay	-0.1241	0.0486	0.0483	0.2194	
Share Medicare	0.0746	0.1900	0.1905	0.3068	+
Log(occupancy rate)	0.4128	0.4635	0.4627	0.5151	+
$\widehat{W}_c y$	0.3181	0.4006	0.4011	0.4801	+
Direct effects					
For Profit	0.0310	0.0816	0.0811	0.1329	+
Chain	0.0114	0.0577	0.0579	0.1040	+
Share Private pay	-0.1270	0.0499	0.0496	0.2254	
Share Medicare	0.0765	0.1951	0.1956	0.3159	+
Log(occupancy rate)	0.4230	0.4758	0.4752	0.5295	+
Indirect effects					
For Profit	0.0180	0.0516	0.0502	0.0924	+
Chain	0.0068	0.0365	0.0356	0.0706	+
Share Private pay	-0.0803	0.0317	0.0306	0.1494	
Share Medicare	0.0459	0.1234	0.1203	0.2174	+
Log(occupancy rate)	0.2040	0.3010	0.2973	0.4168	+

Notes: + indicates  $> 0$  at the 5% level.  $\widehat{W}_c = 0.2W_1 + 0.8W_2$

equals 0.0556, suggesting these weights are not highly correlated with each other.

The model we wish to estimate is still the same, except that  $W_c$  is now constructed as follows:

$$W_c = \gamma_1 W_1 + \gamma_2 W_2 + (1 - \gamma_1 - \gamma_2) W_3 \quad (26)$$

The posterior model probability results from analyzing this 3-way convex combination of connectivity matrices, using a 0.05 grid of values for  $\gamma_1, \gamma_2$  are shown in Table 9.<sup>21</sup> Posterior model probabilities show that models including  $W_3$  receive no posterior probability support, confirming that a model based on the convex combination of the two weight matrices  $W_1, W_2$  has more support from the data. These results are promising since they suggest that we can use the simple grid approach from Section 2.1 that compares models conditional on their connectivity matrices to determine the relevance of weighting matrices in the cross-sectional dependence scheme defined by the convex combination. One issue that often arises when practitioners attempt to include multiple weight matrices in a model is the question of redundancy. Should we add another weight matrix to our model? The appropriate

<sup>21</sup>Again, we only report the convex combinations with non-zero posterior model probability based on 4 decimal digits. We also used the exact value of the log-determinant rather than the approximated value to compute log-marginal likelihood values.

answer would be only if the additional weight matrix reflects a model specification that incorporates another dependence structure for regions/observations that has some posterior model probability support from the data. Our approach allows practitioners to explore this issue.

Table 9: Posterior model probabilities for convex combinations using 3 connectivity matrices

$\gamma_1$	$\gamma_2$	$1 - \gamma_1 - \gamma_2$	Posterior probability	log-marginal Likelihood
0.50	0.50	0	0.0000	-403.3543
0.35	0.60	0.05	0.0002	-401.8628
0.30	0.65	0.05	0.0007	-400.3920
0.45	0.55	0	0.0007	-400.3875
0	0.95	0.05	0.0015	-399.6345
0.25	0.70	0.05	0.0018	-399.4430
0.05	0.90	0.05	0.0024	-399.1896
0.20	0.75	0.05	0.0030	-398.9333
0.10	0.85	0.05	0.0032	-398.8835
0.15	0.80	0.05	0.0036	-398.7744
0.40	0.60	0	0.0072	-398.0717
0	1	0	0.0358	-396.4697
0.35	0.65	0	0.0380	-396.4094
0.05	0.95	0	0.0702	-395.7947
0.30	0.70	0	0.1101	-395.3455
0.10	0.90	0	0.1238	-395.2277
0.15	0.85	0	0.1856	-394.8230
0.25	0.75	0	0.1914	-394.7921
0.20	0.80	0	0.2207	-394.6499

### 5.3 Estimation results based on the MCMC approach

Table 10 shows estimates from the model that uses MCMC estimation to integrate out over the parameter  $\gamma_1$ . In contrast to estimates from the grid approach, we are able to present posterior mean, median and credible intervals for the parameter  $\gamma_1$ . Despite the fact that we present posterior estimates for the parameter  $\gamma_1$ , direct and indirect effects estimates are constructed using all draws for this parameter that arise during MCMC estimation, making the posterior distribution of these effects estimates unconditional on any particular value of the parameter  $\gamma_1$ . This is what distinguishes this approach from the grid approach that produces effects estimates that are conditional on the single estimated value of  $\bar{\gamma}_1 = 0.20$ .

In Table 10, we see a posterior mean for  $\hat{\gamma}_1 = 0.1434$  (and a median of 0.1400), which differs slightly from the the single value of  $\bar{\gamma}_1 = 0.20$  produced by the grid approach to estimation. The similarity between posterior mean and median values indicates symmetry

of the posterior distribution, which reinforces the confidence we have in our results. We note that the estimated  $\gamma_1$  is significantly different from zero, since the 2.5% quantile of its posterior distribution is equal to 0.01. Table 11 compares the results using the two approaches. It reports posterior mean estimates from both models along with their respective credible interval size, computed as the difference between the 97.5% and the 2.5% quantiles values.

The posterior mean estimate for the dependence parameter  $\rho$  is slightly lower in the conditional model (0.34) versus the unconditional model (0.4). However, with the exception of the constant term,  $\beta$  estimates as well as direct and indirect impacts are similar for both models.

For this application, we do not observe significant difference in the size of credible intervals. The grid approach thus performs as well as the MCMC approach, with the difference that the latter allows to perform inference on the convex combination parameters.

Table 10: Posterior parameter estimates based on unconditional MCMC estimates

Variables	Lower 2.5%	Mean	Median	Upper 97.5%	sign/signif
Estimates					
Constant	6.4679	7.3036	7.3007	8.1361	+
For Profit	0.0296	0.0805	0.0806	0.1308	+
Chain	0.0117	0.0575	0.0576	0.1037	+
Share Private pay	-0.1280	0.0480	0.0472	0.2248	
Share Medicare	0.0910	0.2054	0.2038	0.3258	+
Log(occupancy rate)	0.4307	0.4816	0.4816	0.5308	+
$W_c(\gamma_1)y$	0.2661	0.3407	0.3411	0.4161	+
$\gamma_1$	0.0100	0.1434	0.1400	0.3095	
Direct effects					
For Profit	0.0302	0.0822	0.0823	0.1334	+
Chain	0.0119	0.0587	0.0589	0.1061	+
Share Private pay	-0.1307	0.0490	0.0482	0.2302	
Share Medicare	0.0929	0.2097	0.2083	0.3330	+
Log(occupancy rate)	0.4400	0.4917	0.4918	0.5418	+
Indirect effects					
For Profit	0.0143	0.0402	0.0394	0.0712	+
Chain	0.0056	0.0288	0.0279	0.0562	+
Share Private pay	-0.0655	0.0241	0.0229	0.1137	
Share Medicare	0.0429	0.1029	0.1002	0.1800	+
Log(occupancy rate)	0.1662	0.2412	0.2376	0.3347	+

Note: + indicates  $> 0$  at the 5% level.



Table 11: Comparison of unconditional (MCMC) and conditional (grid) estimates ( $\bar{\gamma}_1 = 0.2$ )

Variables	MCMC approach		Grid approach	
	Mean value	Interval size	Mean value	Interval size
Estimates				
Constant	7.3036	1.6681	6.6405	1.7907
For Profit	0.0805	0.1012	0.0795	0.0994
Chain	0.0575	0.0921	0.0562	0.0901
Share Private pay	0.0480	0.3527	0.0486	0.3435
Share Medicare	0.2054	0.2348	0.1900	0.2322
Log(occupancy rate)	0.4816	0.1001	0.4627	0.1023
$W_{cy}$	0.3407	0.1500	0.4006	0.1620
Direct effects				
For Profit	0.0822	0.1032	0.0816	0.1019
Chain	0.0587	0.0942	0.0577	0.0926
Share Private pay	0.0490	0.3609	0.0499	0.3524
Share Medicare	0.2097	0.2402	0.1951	0.2394
Log(occupancy rate)	0.4917	0.1017	0.4758	0.1065
Indirect effects				
For Profit	0.0402	0.0568	0.0516	0.0744
Chain	0.0288	0.0507	0.0365	0.0638
Share Private pay	0.0241	0.1792	0.0317	0.2297
Share Medicare	0.1029	0.1371	0.1234	0.1715
Log(occupancy rate)	0.2412	0.1685	0.3010	0.2128

## 6 Conclusion

We propose two approaches to estimating models based on convex combinations of connectivity matrices, where each matrix can reflect a different type of cross-sectional dependence structure assigned to the sample observations. Our approaches represent Bayesian alternatives to methods set forth in Hazir et al. (2016).

We argue that the convex combination approach has advantages over past spatial econometrics literature that has extended conventional spatial regression models by directly introducing multiple weight matrices. LeSage and Pace (2011) point to a number of estimation and interpretive issues that arise for models that directly introduce multiple weight matrices. Since the convex combination approach relies on a single weight matrix, an advantage is that conventional spatial regression algorithms can be used for estimation and inference.

We discuss several practical concerns that arise for successful use of these models by practitioners, including: the distinction between connectivity matrices and regressors and interpretative issues that arise when a functional relationship exists between these; the

impact of correlation between multiple connectivity matrices and tests that can be used to assess this; and issues that can arise regarding inference at the boundary of the parameter space for  $\gamma$ .

Monte Carlo experiments were carried out to compare our two proposed approaches to estimating these models, which showed that one approach can suffer from bias in average direct and indirect effects as well as incorrect coverage intervals. Incorrect coverage intervals can lead to improper inferences regarding the significance of effects estimates that are typically the focus of spatial regression models.

To illustrate the method, we apply the two proposed approaches to a cross-sectional salary benchmarking model from Blankmeyer et al. (2011) for compensation of CEOs in 856 nursing homes located in the state of Texas. We introduce a convex combination of spatial dependence and peer group cross-sectional dependence, and find that peer-group dependence is relatively more important (83%) in determining CEO compensation than spatial dependence (17%), with both types of dependence playing a role in explaining variation of CEO compensation across the sample of nursing homes.

A future extension of the methods proposed here would involve static panel data models. The grid approach to estimating the convex combination models rely on Bayesian methods described in LeSage (2015) to calculate log-marginal likelihoods (and associated posterior model probabilities) for both cross-sectional and static panel data spatial regression models, which should make this type of extension simple. A related point is that the Bayesian model comparison methods described in LeSage (2014) focus on calculating posterior probabilities that compare both cross-sectional and panel data SDM, SDEM, SLX specifications. This should make it possible to embed exploration of convex combinations of weight matrices in a broader search for the model specification.

A drawback to implementation of the MCMC approach used here is reliance on pre-calculation of the log-determinant term that appears in the conditional distributions for the parameters  $\rho$  and  $\gamma$  over a grid of values for these parameters. While this works effectively to avoid calculation of log-determinants during each pass of the MCMC sampler for the case of two or three weight matrices, it does not scale well to cases involving more weight matrices. A subject for future research is to develop a more computationally efficient approach for extending the MCMC approach set forth here to problems involving more weight matrices.

## Acknowledgments

Nicolas Debarsy acknowledges financial support of the Region Centre Val de Loire (APRIA 2014 00094514). We would also like to thank Sergio Rey, Ingmar Prucha, Xi Qu and participants of the Department of Economics of Jiao Tong University 16<sup>th</sup> International Workshop in Spatial Econometrics and Statistics (France), 1<sup>st</sup> International Conference in Econometrics and Statistics (Honk-Kong), Western Regional Science Association 56<sup>th</sup> Annual Meeting(Mexico), 61<sup>th</sup>World Statistic Congress (Morocco).

## References

- Anderson, J.E., van Wincoop E., 2003, Gravity with gravitas: A solution to the border puzzle *American Economic Review*, 93, 170-192.
- Anderson, J.E., van Wincoop E., 2004, Trade costs *Journal of Economic Literature*, 42, 691-751.
- Anselin L., 1988, *Spatial Econometrics - Methods and Models*, Kluwer Academic Publishers, Dordrecht.
- Badinger, H., P. Egger, 2011, Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances *Papers in Regional Science*, 90(1), 213–235.
- 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.
- Behrens, K.,C. Ertur, W. Koch, 2012, Dual gravity : Using spatial econometrics to control for multilateral resistance *Journal of Applied Econometrics*, 27(5), 773–794.
- 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.
- Bramoullé Y., H. Djebbari, B. Fortin, 2009, Identification of peer effects through social networks, *Journal of Econometrics* 150, 41–55.
- Burridge P., 2012, Improving the J test in the SARAR model by likelihood-based estimation, *Spatial Economic Analysis*, 7, 75–107.

- Burridge, P., B. Fingleton, 2010, Bootstrap inference in spatial econometrics: the J-test, *Spatial Economic Analysis*, 5, 93-119
- Corrado, L., B. Fingleton, 2012, Where is the economics in spatial econometrics? *Journal of Regional Science*, 52, 210–239.
- Cox D. R., 1961, Tests of separate families of hypotheses, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 1, 105–123.
- Cox D. R., 1962, Further results on tests of separate families of hypothesis, *Journal of the Royal Statistical Society, Serie B*, 24, 406–424.
- Davidson R., J. G. MacKinnon, 1981, Several tests for model specification in the presence of alternative hypotheses, *Econometrica*, 49, 781–793.
- Debarsy N., C. Dossougoin, C. Ertur, J-Y. Gnabo, 2016, Measuring sovereign risk spillovers and assessing the role of transmission channels: A spatial econometrics approach, CORE Discussion Papers - 2017/53.
- Debarsy N., C. Ertur, 2016, Interaction matrix selection in spatial econometrics with an application to growth theory, Working Paper, DR LEO 2016-05.
- Elhorst, J.P., D.J. Lacombe, G. Piras, 2011, On model specification and parameter space definitions in higher order spatial econometric models, *Regional Science and Urban Economics*, 42, 211–220.
- Jin F., L.-F. Lee, 2013, Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances, *Regional Science and Urban Economics*, 43, 590–616.
- 'eniaux G., D. Martinetti, 2017, A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models, *Regional Science and Urban Economics*, Forthcoming.
- Geweke, J., 1988, The secular and cyclical behavior of real GDP in 19 OECD countries, 1957-1983, *Journal of Business & Economic Statistics*, 6, 479–486.
- Ghigolino C., S. Goyal, 2010, Keeping up with the Neighbors: Social Interaction in a Market Economy, *Journal of the European Economic Association*, 8, 90–119.

- Hagemann A., 2012, A simple test for regression specification with non-nested alternatives, *Journal of Econometrics*, 166, 247–254.
- Hazir, C.S., J.P. LeSage, C. Autant-Bernard, 2016, The role of R&D collaboration networks on regional innovation performance, *Papers in Regional Science*, forthcoming.
- Holloway, G., B. Shankara, S. Rahman, 2002, Bayesian spatial probit estimation: a primer and an application to HYV rice adoption, *Agricultural Economics*, 27(3), 383–402.
- Kelejian H. H., 2008, A spatial J-test for model specification against a single or a Set of Non-Nested Alternatives, *Letters in Spatial and Resource Sciences*, 1, 3–11.
- Kelejian H. H., I. R. Prucha, 2010, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics*, 157, 53–67
- Kelejian H. H., G. Piras, 2011, An extension of Kelejian’s J-test for non-nested spatial models, *Regional Science and Urban Economics*, 41, 281–292.
- Koch, W., J.P. LeSage, 2015, Latent multilateral trade resistance indices: theory and evidence, *Scottish Journal of Political Economy*, 62(3) 264–290.
- Lacombe, D.J., 2004, Does econometric methodology matter? An analysis of public policy using spatial econometric techniques. *Geographical Analysis*, 36, 105–118.
- Lee L.-F., 2007, Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* 140, 333–374.
- Lee, L.-F., X. Liu, 2010, Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances, *Econometric Theory* 26, 187–230.
- Lee L.-F., J. Yu, 2012, QML estimation of spatial dynamic panel data models with time varying spatial weights matrices, *Spatial Economic Analysis*, 7, 31–74
- Lee L.-F., J. Yu, 2014, Efficient GMM estimation of spatial Dynamic panel data models with fixed effects, *Journal of Econometrics*, 180, 174–197.
- LeSage, J. P., R. K. Pace, 2009, *Introduction to Spatial Econometrics*, (Boca Raton: Taylor Francis/CRC Press).

- LeSage J. P., 2014, Spatial econometric panel data model specification: A Bayesian approach, *Spatial Statistics*, 9, 122–145.
- LeSage, J. P., 2015, Software for Bayesian cross section and panel spatial model comparison, *Journal of Geographical Systems*, 17, 297–310.
- 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.
- Liu X., E. Patacchini, Y. Zenou, 2014, Endogenous peer effects: local aggregate of global average?, *Journal of Economic Behavior and Organization*, 103, 39–59.
- Nicholls, D.F., A.R. Pagan and R.D. Terrell, 1975, The estimation and use of models with moving average disturbance terms: a survey, *International Economic Review*, 16, 112–134.
- Ord K., 1975, Estimation methods for models of spatial interaction, *Journal of the American Statistical Association*, 70, 120–126.
- Pace, R.K., R.P. Barry, 1997, Quick computation of spatial autoregressive estimators, *Geographical Analysis*, 29, 232–246.
- Pace, R.K., J.P. LeSage, 2002, Semiparametric maximum likelihood estimates of spatial dependence, *Geographical Analysis*, 34(1), 75–90.
- Pace, R.K., J.P. LeSage and S. Zhu 2012, Spatial Dependence in Regressors, *Advances in Econometrics* Volume 30, Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano and Eric Hillebrand (Series Eds., Volume editors: Dek Terrell and Daniel Millimet) , Emerald Group Publishing Limited, pp. 257–295.
- Patacchini E., Y. Zenou, 2012, Juvenile delinquency and conformism, *The Journal of Law, Economics and Organization*, 28, 1–31.
- Qu X., L.-f. Lee, 2015, Estimating a spatial autoregressive model with an endogenous spatial weight matrix, *Journal of Econometrics*, 184, 209–232.

Smith, T.E., J.P. LeSage, 2004, A bayesian probit model with spatial dependencies, in *Advances in Econometrics: Volume 18: Spatial and Spatiotemporal Econometrics*, (Oxford: Elsevier Ltd), James P. LeSage and R. Kelley Pace (eds.), pp. 127–160.

## A Appendix

### A.1 Conditional distributions for alternative spatial regression specifications involving a convex combination of weights

We set forth the conditional distributions required for MCMC estimation of alternatives to the SAR spatial regression specification described in the text. These include: the spatial Durbin model (SDM), spatial lag of  $X$  model (SLX) and spatial Durbin error model (SDEM). LeSage (2014) argues these three alternative spatial regression specifications should be used in most applied work.

The griddy Gibbs procedures used to sample the parameters  $\rho$  and  $\gamma$  are also set forth.

#### A.1.1 The SDM model

The SDM specification in (27) involves both  $X$  and  $W_c(\gamma)X$  as matrices of explanatory variables, where  $\gamma = [\gamma_1, \dots, \gamma_L]$  denotes the vector containing all the parameters of the convex combination. We can collect the set of explanatory variables in the matrix  $Z = [I_n, X, W_cX]$ , with associated coefficients vector  $\psi = [\alpha, \beta', \theta']'$ . Rewriting (27) using  $Z$  and  $\psi$  allows to express both the SAR the SDM model using the same matrix expressions. Expressions for the conditional distributions required for MCMC sampling of the SAR specification can be extended to the case of the SDM specification by replacing  $X$  with  $Z$  and  $\beta$  by  $\psi$  in expressions (11) to (17).

$$\text{SDM} : y = \rho W_c(\gamma)y + X\beta + W_c(\gamma)X\theta + \varepsilon \quad (27)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$W_c(\gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$$

#### A.1.2 The SDEM model

For the SDEM model in (28), we rely on the same normal prior for  $\psi = (\beta', \theta')'$ , and the uniform prior for  $\lambda$ , since this scalar dependence parameter is constrained to lie in the open interval:  $(-1, 1)$ , which can be imposed during MCMC estimation using rejection sampling. Similarly, we simply impose the closed interval  $[0, 1]$  for each  $\gamma_\ell$ ,  $\ell = 1, \dots, L - 1$  during



MCMC estimation and ensure that  $\sum_{\ell=1}^L \gamma_\ell = 1$  using  $\gamma_L = (1 - \sum_{\ell=1}^{L-1} \gamma_\ell)$ . The inverse Gamma( $\bar{a}, \bar{b}$ ) distribution is used for the parameter  $\sigma^2$  with  $\bar{a}, \bar{b} \rightarrow 0$ , producing little prior information.

$$\text{SDEM} : y = X\beta + W_c(\gamma)X\theta + (I_n - \lambda W_c(\gamma))^{-1}u. \quad (28)$$

$$u \sim N(0, \sigma^2 I_n)$$

$$W_c(\gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$$

The conditional posterior for  $\psi$  (given  $\lambda, \gamma, \sigma^2$ ) takes the form in (29).

$$p(\psi | \lambda, \gamma, \sigma^2) \sim N(\psi^*, \Sigma^*) \quad (29)$$

$$\psi^* = \Sigma^* (Z^{*\prime} y^* + \sigma^2 \bar{\Sigma}_\psi^{-1} \bar{\psi})$$

$$\Sigma^* = (Z^{*\prime} Z^* + \sigma^2 \bar{\Sigma}_\psi^{-1})^{-1}$$

$$y^* = y - \lambda W_c(\gamma) y \quad (30)$$

$$Z^* = Z - \lambda W_c(\gamma) Z \quad (31)$$

$$Z = [\iota_n, X, W_c(\gamma) X]$$

$$W_c(\gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$$

The conditional posterior for  $\sigma^2$  (given  $\psi, \gamma, \lambda$ ) takes the form in (32), when we set the prior parameters  $\bar{a} = \bar{b} = 0$ . The vectors  $y^*, Z^*$  are as defined in (30) and (31) above.

$$\begin{aligned} p(\sigma^2 | \psi, \lambda, \gamma) &\propto (\sigma^2)^{-\left(\frac{n}{2}\right)} \exp\left(-\frac{1}{2\sigma^2} (y^* - Z^* \psi)' (y^* - Z^* \psi)\right) \\ &\sim IG(a_1, b_1) \end{aligned} \quad (32)$$

$$a_1 = n/2$$

$$b_1 = (y^* - Z^* \psi)' (y^* - Z^* \psi) / 2$$

The (log) conditional posterior for  $\lambda$  (given  $\psi, \gamma, \sigma^2$ ) takes the form in (33), where we use

the expression:  $(e'e)(\lambda, \gamma)$  to indicate that the sum of squared errors depends on parameters  $\lambda, \gamma$ .

$$\begin{aligned}
p(\lambda|\psi, \gamma, \sigma^2) &\propto -\frac{n}{2}\ln\sigma^2 + \ln|I_n - \lambda W_c(\gamma)| \\
&\quad - \frac{1}{2\sigma^2}(e'e)(\lambda, \gamma) \\
e'e(\lambda, \gamma) &= ((y^* - Z^*\psi)'(y^* - Z^*\psi))
\end{aligned} \tag{33}$$

As in the case of  $\rho$  for the SDM specification, this distribution is not one of known form. We sample the parameter  $\lambda$  from this conditional distribution using the same griddy Gibbs approach described for  $\rho$  in Appendix A.2.1.

The (log) conditional posterior for  $\gamma_\ell$  (given  $\psi, \lambda, \sigma^2, \gamma_{-\ell}$ ) takes an unknown form as in the case of  $\lambda$ , so we rely on griddy Gibbs sampling for these parameters. After sampling an update for all the parameter vector  $\gamma$ , the matrix  $W_c(\gamma) = \sum_{\ell=1}^L \gamma_\ell W_\ell$  is reformulated using the most recently updated value.

$$\begin{aligned}
p(\gamma_\ell|\psi, \lambda, \sigma^2, \gamma_{-\ell}) &\propto -\frac{n}{2}\ln\sigma^2 + \ln|I_n - \lambda W_c(\gamma_\ell|\gamma_{-\ell})| \\
&\quad - \frac{1}{2\sigma^2}e'e(\lambda, \gamma_\ell, \gamma_{-\ell})
\end{aligned} \tag{34}$$

$$e'e(\lambda, \gamma_\ell, \gamma_{-\ell}) = ((y^* - Z^*\psi)'(y^* - Z^*\psi)) \tag{35}$$

$$W_c(\gamma_\ell|\gamma_{-\ell}) = \gamma_\ell W_\ell + \sum_{o \neq \ell} \gamma_o W_o$$

### A.1.3 The SLX model

For the case of the SLX model in (36), we have both  $X$  and  $W_c(\gamma)X$  as matrices of explanatory variables, which we can collect in  $Z = [I_n, X, W_c(\gamma)X]$ , with associated coefficients vector  $\psi = [\alpha, \beta', \theta']'$ . Replacing the matrix  $X$  with  $Z$  and the coefficients vector  $\beta$  with  $\psi$  in the expressions for the conditional distributions of the SAR specification produces conditional distributions required for MCMC sampling of the parameters  $\psi$  and  $\sigma^2$  for the SLX model. We can employ Gibbs sampling from the conditional distributions for  $\psi, \sigma^2$  which take known distributional forms, (multi-normal and inverse Gamma), and rely on griddy Gibbs sampling for the parameters  $\gamma$ . Further, relying on the  $Z$  and  $\psi$  notations,

we observe that the SLX is nested by the SAR, when the parameter  $\rho$  is set to zero. Hence, to estimate the SLX, we use the SAR framework without sampling for  $\rho$ , which is fixed at a zero value.

$$\begin{aligned}
\text{SLX : } y &= \alpha \iota_n + X\beta + W_c(\gamma)X\theta + \varepsilon & (36) \\
\varepsilon &\sim N(0, \sigma^2 I_n) \\
W_c(\gamma) &= \sum_{\ell=1}^L \gamma_\ell W_\ell
\end{aligned}$$

## A.2 Griddy Gibbs sampling

### A.2.1 Griddy Gibbs sampling for $\rho$ ( and/or $\lambda$ )

For griddy Gibbs sampling we rely on expressing the log conditional posterior distribution for the parameter  $\rho$  as a vector over a grid of values for  $\rho$ . A restriction on the grid that  $-1 < \rho < 1$  is imposed to ensure stability of the spatial autoregressive process to this range.<sup>22</sup> This vector over a grid of values for  $\rho$  allows us to use univariate numerical integration to obtain a normalizing constant and then construct a cumulative density function (CDF) for the conditional posterior for the parameter  $\rho$ . Given this CDF, we produce a draw from the conditional distribution using inversion, an approach introduced by Smith and LeSage (2004).

Specifically, we can express the SAR conditional distribution for  $\rho$  as:

$$\begin{aligned}
p(\rho|\beta, \gamma, \sigma^2) &\propto -\frac{N}{2} \ln \sigma^2 + \ln |I_n - \rho W_c(\gamma)| \\
&\quad - \frac{1}{2\sigma^2} s^2(\rho|\gamma) & (37)
\end{aligned}$$

$$s^2(\rho|\gamma) = ([I_n - \rho W_c(\gamma)]y - X\beta)' ([I_n - \rho W_c(\gamma)]y - X\beta)$$

We draw on the vectorization scheme for the grid of  $q$  values for  $\rho$  from Pace and Barry (1997), to produce the following vector  $q \times 1$  vector over values of  $\rho$ , which is conditional on a particular set of values for the parameter vector  $\gamma$ , indicated using:  $s^2(\rho|\gamma)$ .

---

<sup>22</sup>A similar restriction is imposed on the parameter  $\lambda$ .

$$\begin{pmatrix} \text{Ln } p(\rho_1|y) \\ \text{Ln } p(\rho_2|y) \\ \vdots \\ \text{Ln } p(\rho_q|y) \end{pmatrix} \propto \begin{pmatrix} \text{Ln}|I_n - \rho_1 W_c(\gamma)| \\ \text{Ln}|I_n - \rho_2 W_c(\gamma)| \\ \vdots \\ \text{Ln}|I_n - \rho_q W_c(\gamma)| \end{pmatrix} - \left(\frac{n-k}{2}\right) \begin{pmatrix} \text{Ln}(s^2(\rho_1|\gamma)) \\ \text{Ln}(s^2(\rho_2|\gamma)) \\ \vdots \\ \text{Ln}(s^2(\rho_q|\gamma)) \end{pmatrix} \quad (38)$$

and we note that the sum of squared errors terms  $s^2(\rho_i|\gamma)$  can be expressed as a vector over values  $\rho_i, i = 1, \dots, q$  using:

$$s^2(\rho_i|\gamma) = e_o' e_o - 2\rho_i e_d'(\gamma) e_o + \rho_i^2 e_d'(\gamma) e_d(\gamma) \quad (39)$$

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

$$\beta_d(\gamma) = (X'X)^{-1} X'W_c(\gamma)y \quad (40)$$

$$e_o = y - X\beta_o$$

$$e_d(\gamma) = W_c(\gamma)y - X\beta_d(\gamma)$$

This vector  $s^2(\rho_i|\gamma)$  can be calculated rapidly for a given value of the parameter vector  $\gamma$ . We also need the log-determinant term  $\text{Ln}|I_n - \rho_i W_c(\gamma)|$  over the grid of values for  $\rho_i$ . The log-determinant terms are pre-calculated over grids of values for  $\rho, \gamma_\ell, \ell = 1, \dots, L$  and are stored in a 3-dimensional matrix of size  $(np \times ng \times L)$ , where  $np$  is the number of values for  $\rho_i$  in the grid,  $ng$  the number of values for each  $\gamma_\ell$ , and  $L$  is the size of the parameter vector  $\gamma$ . During MCMC sampling, we lookup the appropriate vector of log-determinant values based on the current values of  $\gamma$ .

### A.2.2 Griddy Gibbs sampling for $\gamma_\ell, \ell = 1, \dots, L$

A similar approach to that used for the parameter  $\rho$  can be used for sampling  $\gamma_\ell, \ell = 1, \dots, L$  values.

The (log) conditional posterior for  $\gamma_\ell$  in the case of the SAR model (given  $\beta, \rho, \sigma^2, \gamma_{-\ell}$ ) can be written as in (41), where we use  $W_c(\gamma_\ell|\gamma_{-\ell})$  to indicate that we are sampling each  $\gamma_\ell$  conditional on all other parameters of the convex combination.

$$p(\gamma_\ell|\beta, \rho, \sigma^2, \gamma_{-\ell}) \propto -\frac{n}{2} \ln \sigma^2 + \ln |I_n - \rho W_c(\gamma_\ell|\gamma_{-\ell})| - \frac{1}{2\sigma^2} s^2(\gamma_\ell|\rho, \gamma_{-\ell}) \quad (41)$$

We want to express this as a vector over a grid of values for  $\gamma_\ell$  that we designate using  $\gamma_\ell^i, i = 1, \dots, nq$ , where  $nq$  is determined by the grid size. The sum of squared error terms can be expressed as shown in (42).

$$\begin{aligned}
s^2(\gamma_\ell^i, |\rho, \gamma_{-\ell}) &= e_o' e_o - 2\rho e_d' e_o(\gamma_\ell^i | \gamma_{-\ell}) + \rho^2 e_d'(\gamma_\ell^i | \gamma_{-\ell}) e_d(\gamma_\ell^i | \gamma_{-\ell}) & (42) \\
e_o &= y - X\beta_o \\
e_d(\gamma_\ell^i | \gamma_{-\ell}) &= W_c^i(\gamma_\ell^i | \gamma_{-\ell}) y - X\beta_d \\
\beta_o &= (X'X)^{-1} X'y \\
\beta_d &= (X'X)^{-1} X'W_c^i(\gamma_\ell^i | \gamma_{-\ell}) y & (43) \\
W_c^i(\gamma_\ell^i | \gamma_{-\ell}) &= \gamma_\ell^i W_i + \sum_{o \neq i} \gamma_o W_o
\end{aligned}$$

This vector  $s^2(\gamma_\ell^i | \gamma_{-\ell}, \rho)$  can be calculated rapidly for a given value of  $\gamma_\ell$  and  $\rho$  conditional on the other values  $\gamma_{-\ell}$  during MCMC sampling. The log-determinant terms needed for this log conditional posterior distribution are extracted from the pre-calculated grids of values for  $\rho, \gamma_\ell, \ell = 1, \dots, L$  using lookup here as well.