Is a matrix exponential specification suitable for the modeling of spatial correlation structures?

Magdalena E. Strauß[a,*], Maura Mezzetti[b,*], Samantha Leorato[b]

[a] *MRC Biostatistics Unit, University of Cambridge, Robinson Way, Cambridge CB2 0SR, UK*
[b] *Department of Economics and Finance, Università Tor Vergata, Via Columbia 2, 00133 Rome, Italy*

Abstract

This paper investigates the adequacy of the matrix exponential spatial specifications (MESS) as an alternative to the widely used spatial autoregressive models (SAR). To provide as complete a picture as possible, we extend the analysis to all the main spatial models governed by matrix exponentials comparing them with their spatial autoregressive counterparts.

We propose a new implementation of Bayesian parameter estimation for the MESS model with vague prior distributions, which is shown to be precise and computationally efficient. Our implementations also account for spatially lagged regressors. We further allow for location-specific heterogeneity, which we model by including spatial splines. We conclude by comparing the performances of the different model specifications in applications to a real data set and by running simulations. Both the applications and the simulations suggest that the spatial splines are a flexible and efficient way to account for spatial heterogeneities governed by unknown mechanisms.

Key words: Matrix exponential, Covariance matrix, Spatial correlation

---

1. Introduction

Data collected from geographic areas such as countries, regions, states, or individual points in space often exhibit spatial dependence, and require specific estimation

*Corresponding author

Email addresses: magdalena.strauss@mrc-bsu.cam.ac.uk (Magdalena E. Strauß), maura.mezzetti@uniroma2.it (Maura Mezzetti), samantha.leorato@uniroma2.it (Samantha Leorato)

methods to account for the lack of independence among the data. In recent years the economics literature has seen an increasing number of theoretical and applied econometric studies involving spatial dependence. While the interest in spatial models in economics is relatively recent, spatial models have a long history in the regional science, epidemiology and geography literature (see Anselin & Florax [1] for detailed references).

The widely used spatial autoregressive (SAR) approach was first introduced by Whittle [34] and refers to the autoregressions that occur simultaneously at each data location. One drawback of the SAR model is that it requires specialized techniques for large samples. Ciu et al. [4] first proposed exponential operators to specify a covariance matrix and also pointed out some advantages of the matrix exponential, but focused on general (non-spatial) covariance matrices. Later, LeSage & Pace [20] proposed to apply the matrix exponential specification in a spatial context, as an alternative to the widely used SAR model. The resulting matrix exponential spatial specification model (MESS) replaces the conventional geometric decay of influence over space with an exponential pattern of decay. The MESS model has advantages, relative to the SAR, deriving from the characteristics of the matrix exponential reviewed in Section 2. However, it also has some disadvantages, the first of which is the difficult interpretation of the *correlation* parameter, which was also noted by LeSage & Pace [20]. One further concern related to the use of the MESS model, was raised by Rodrigues et al. [24], who recently showed that it often induces opposite signs for the marginal and conditional correlations between two areas. We briefly discuss these two aspects in the Appendix, which we devote to a comparison between MESS and SAR marginal effects and covariance structures.

Matrix exponentials can be introduced to define either the interaction between dependent variables or the spatial covariance of the errors. To these different approaches correspond two subclasses of the MESS models, usually referred to as MESS models and MESS error models, respectively. These models are alternatives to the SAR and spatial error models (SEM). The final goal of this paper is to contribute to the literature on the matrix exponential model, by assessing its validity, both on its own and relative to its main competitor, the SAR model. To take up this challenge and to allow a wider comparison with the SAR models, in Section 3 the different specifications of the spatial models with matrix exponential covariance are illustrated. As a possible way to allow MESS error models to account for location-specific heterogeneity, in Section 3.2, we moreover explore the effects of introducing spatial splines to cope with uncertainty of the spatial structure, which is acknowledged to be one common weak point of spatial linear regression models. We focus in particular on Bayesian estimation of the models: in Section 4 we propose a new implementation of Bayesian parameter estimation with vague prior distributions for both MESS and MESS error models.

To our knowledge, Bayesian approaches have never been used for the estimation of the latter. In fact, ever since the work of LeSage & Pace [20], the literature has mainly focused on the first class, and MESS error models have been neglected, except for a brief description in LeSage & Pace [21].

In contrast to previous model implementations, our method does not use Taylor series expansion with a fixed number of terms to approximate the matrix exponential; using an appropriate R package, the method used in our approach ensures that our approximation to the matrix exponential is always within a given fixed small interval around the true value. For the MESS model, we use an algorithm based on Krylov subspaces techniques developed by Sidje & Stewart [27]. Like the Taylor expansion method discussed in [21], Sidje & Stewart [27]'s algorithm directly computes the action of a matrix exponential on a vector without computing the matrix exponential itself. It was shown to be very efficient in Sidje & Stewart [27], and it avoids lacking control on roundoff errors that may occur in Taylor series approximations due to alternating signs of the terms in the series. Because of these differences, a reasonable comparison between our implementation with those based on Taylor approximation of the exponential series, as the one proposed by [21], is in terms of computation time. This is presented in Section 5.6.

Moreover in Section 5, by applying the different MESS models in an econometric application and in simulated data, we could assess its predictive ability with different weight matrices, and we find comparable performances relative to the SAR model with the same weight matrix choices.

The application shows that the model with splines outperforms most of its competitors in terms of predictive accuracy. Moreover, a simulation shows that it is more robust to model misspecification than the MESS model . This suggests that the model with splines could be a promising development of both the MESS and SAR models. In particular, we argue that the extension of spatial regression models through the introduction of splines is able to mitigate the possible misspecification of the spatial weight structure. Finally, Section 6 offers some concluding remarks.

2. Matrix exponential

Matrix exponentials have been used as the basis of covariance structures by several authors [17, 4, 6, 12, 23, 18], because of some particular properties of the exponential operator. Some of these models (such as [17], [4]) use the matrix exponential to model general non-spatial covariance structures.

The matrix exponential of an $n \times n$ matrix $\mathbf{A}$, defined as

$$\mathbf{C} = \exp(\mathbf{A}) = \sum_{j=0}^{\infty} \frac{\mathbf{A}^s}{s!} \tag{1}$$

has a number of properties which make it suitable to model covariance matrices. We restrict the following list to those properties that are relevant to the spatial model discussed in this paper. For a complete list of properties relevant to other, non-spatial models, see Ciu et al. [4].

(a) For any square matrix $\mathbf{A}$, $(\exp(\mathbf{A}))^{-1} = \exp(-\mathbf{A})$. This can be seen from the Taylor series expansion (1).

(b) The logarithm of the determinant is $\log|\mathbf{C}| = \mathrm{tr}(\mathbf{A})$.

It should be noted that property (a) implies, in particular, that $\exp(\mathbf{A})$ is not singular for any matrix $\mathbf{A}$, since the matrix exponential of real valued matrices always leads to positive definite covariance matrices, thus eliminating the need to restrict the parameter space, or to test for positive definiteness during optimization. Property (b) is particularly relevant to LeSage & Pace's model ([20]), as they choose matrices $\mathbf{A}$ with $\mathrm{trace}(\mathbf{A}) = 0$. Therefore, the MESS model does not require the computation of determinants appearing in the log-likelihood. Moreover, from (a), the inversion of the matrix exponential takes a simple mathematical form that is easy to implement in applied practice, and together with (b), the use of the matrix exponential covariance leads to a log-likelihood where a troublesome term involving the log determinant vanishes.

3. MESS models for spatially correlated data

After this general introduction to the matrix exponential, we now present the MESS model, first introduced by [20], as an alternative to spatial autoregressive models in the dependent variable. Further, we present different specifications of spatial models with matrix exponential covariance.

3.1. MESS model

In the MESS model originally proposed by LeSage & Pace [20], a matrix exponential is used to model the spatial interaction between dependent variables. The basic model is specified as follows:

$$S(\rho)\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad\qquad S(\rho) = \exp(\rho\mathbf{D}) = \sum_{j=0}^{\infty} \frac{\rho^j \mathbf{D}^j}{j!} \tag{2}$$

where $\mathbf{X}$ is a $n \times K$ matrix of covariates, β a vector of coefficients, $\mathbf{D}$ is a spatial weight matrix, $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I_n})$, and ρ is a scalar parameter reflecting the level of spatial interaction. The model may be rewritten in the following way:

$$\mathbf{y} = S^{-1}(\rho)\mathbf{X}\beta + \nu. \tag{3}$$

As $S^{-1}(\rho) = \exp(-\rho\mathbf{D})$, and because of the normality of ε, the covariance matrix of ν is

$$\Sigma_\nu = \sigma^2 \exp(-\rho\mathbf{D}) \exp(-\rho\mathbf{D}'). \tag{4}$$

As we mentioned in Section 2, $\exp(\rho\mathbf{D})$ is not singular, even if $\mathbf{D}$ is. This is an advantage of LeSage & Pace [20]'s model as compared to the SAR spatial autoregressive model (which corresponds to $S(\rho) = \mathbf{I} - \rho\mathbf{D}$), because there is no need to impose restrictions on either the parameter or on $\mathbf{D}$ to have a well defined reduced form. This means that ρ may assume any value on the real line without causing the covariance matrix to be singular. It has been pointed out that, contrary to the SAR model, that might become unstable in case of strong spatial correlation, the MESS model remains stable independently on the value of ρ. Debarsy et al. [5], in particular, suggest that this feature makes the MESS model particularly suited to observed data that do not show unstable behaviors. On the other hand, for row standardized weight matrices the parameter λ of the SAR model is usually confined in $(-1,1)$ and this implies a natural interpretation of λ as a spatial correlation parameter, that is completely lost for the MESS model.

A possible generalization of the MESS model can be obtained by the introduction of spatially lagged regressors, thus defining a MESS Durbin model:

$$\exp(\rho\mathbf{D})\mathbf{y} = \mathbf{X}\beta + \mathbf{D}\mathbf{X}\theta + \varepsilon \tag{5}$$

where θ is a vector of coefficients. Note that if $\mathbf{X}$ includes the intercept and $\mathbf{D}$ is row standardized, then $\mathbf{D}\mathbf{X}\theta$ should be replaced by $\mathbf{D}\mathbf{X}_2\theta$, where $\mathbf{X} = [\iota, \mathbf{X}_2]$, to ensure full rank; then θ has size $K-1$.

Model (5) was considered by Piribauer & Fischer [23], who proposed a Bayesian model averaging approach, to deal with the uncertainty of the spatial structure.

3.2. MESS error model

The matrix exponential can be used to specify the autocorrelation structure of the dependent variables or of the errors. The latter case gives the MESS error model, briefly introduced by [21]:

$$\mathbf{y} = \mathbf{X}\beta + \nu \tag{6}$$

where ν has covariance matrix (4)[1].

Clearly, also in this case, we can allow for spatially lagged covariates to be included in the equations

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{D}\mathbf{X}\theta + \nu. \tag{7}$$

[1]Our formulation slightly differs from the MESS spatial error defined in [21], where $\Sigma_\nu = \sigma^2 \exp(\lambda\mathbf{D})$, and $\mathbf{D}$ must be symmetric.

One of the limitations of error models is that they do not allow for heterogeneous effects. While the Durbin version of MESS error model allows for spillover effects of neighboring locations, variations of a regressor occurring at a given location have a homogeneous effect on the dependent variable of the same location (see Table 1 below). The possibility to account for location-specific effects in a flexible way motivates the generalization of the MESS error model by the introduction of spatial splines. The model we propose is more flexible and, compared to the original MESS error model, shows better predictive ability in our application to house prices presented in Section 5 as well as in the simulations. While the residuals are assumed to be correlated according to the MESS model, the outcomes are additionally assumed to follow a spatial trend modeled by means of splines of the coordinates of the centroids of the regions.

We specify the MESS error model with spatial splines as follows:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{X}^{*}\beta^{*} + \exp(-\rho\mathbf{D})\varepsilon \tag{8}$$

where $\mathbf{X}^{*}$ is the basis matrix of natural splines of the coordinates of the centroids of the regions of the lattice.

The MESS model with spatial splines accounts for heterogeneity in the location, but does not allow for spillover effect; thus, it can be either alternative or complementary to the inclusion of spatially lagged regressors.

In principle, the spatial splines component could also be included easily in the MESS model (3). However, since model (3) intrinsically accounts for spatial heterogeneity of the effects, in this paper we decided to limit the implementation of the spline specification to the error models only.

It should be noted that, as we are using the basis matrix of the splines, estimating the parameters of model (8) is formally equivalent to estimating the MESS error model (6).

3.3. Impact measures

In general, in spatial models, the β parameters alone are not able to explain the effect of the covariates on the dependent variables. It has been observed that a valid

Model	*Direct effect*$= \partial y_i / \partial x_{ik}$	*Indirect effect*$= \partial y_i / \partial x_{jk}$
MESS	$a(i,i)\beta_k$	$a(i,j)\beta_k$
MESS Durbin	$a(i,i)\beta_j + \sum_{m=1}^{n} a(i,m)d(m,i)\theta_k$	$a(i,j)\beta_k + \sum_{m=1}^{n} a(i,m)d(m,j)\theta_k$
MESS error	β_k	0
MESS error Durbin	β_k	$d(i,j)\theta_k$

Table 1: Direct and indirect effects of the MESS models, $e^{-\rho \mathbf{D}} = \{a(i,j)\}_{ij}$

basis for the identification of spatial spillovers results from a partial derivative interpretation of the impact from changes to the covariates. Thus, the impact of the $k-$th regressor on the dependent variable at the $i-$th location, y_i, is defined as $\frac{\partial y_i}{\partial \mathbf{x}_k} = \left(\frac{\partial y_i}{\partial x_{1k}}, \ldots, \frac{\partial y_i}{\partial x_{ik}}, \ldots, \frac{\partial y_i}{\partial x_{nk}}\right)$. In particular, the effect on y_i of variations of the $k-$th regressor in the $i-$th location, $\frac{\partial y_i}{\partial x_{ik}}$, is called a *direct effect*, while each of the terms $\frac{\partial y_i}{\partial x_{jk}}$, $j \neq i$ is an *indirect effect*. Then, when comparing different models, it is imperative to add impact measures based on partial derivatives.

In this subsection we present the different impact measures related to all the model specifications previously defined. We denote, for convenience, by $a(i,j)$ the $(i,j)-$th element of $e^{-\rho \mathbf{D}}$ and by $d(i,j)$ the $(i,j)-$th component of $\mathbf{D}$.

Note that the direct and indirect effects of the error models do not differ from the corresponding effects of SAR error models (SEM). The direct effect in the Durbin model is constant and equal to β_k, provided we assume that the matrix $\mathbf{D}$ has zero elements in the main diagonal. The MESS error model, with or without splines, does not account for spillover effects; the splines only are able to capture location-specific heterogeneities.

Following LeSage & Pace [21], a summary indicator for the direct effect is given by the average of the diagonal elements of the matrix as outlined in Table 1, and a summary indicator for the indirect effect is the average of either the row sums or the column sums of the off-diagonal elements of that matrix, thus

$$\bar{M}_{dir}(k) = n^{-1} \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_{ik}}; \quad \bar{M}_{tot}(k) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_{jk}}; \quad \bar{M}_{in}(k) = \bar{M}_{tot}(k) - \bar{M}_{dir}(k). \tag{9}$$

The average row effect represents the impact on a particular element of the dependent variable as a result of a unit change in all elements of an independent variable, while the average column effect represents the impact of changing a particular element of an independent variable on the dependent variable of all other units. However, since the numerical magnitudes of these two calculations of the indirect effect are the same, it does not matter which one is used. Generally, the indirect effect is interpreted as the impact of changing a particular element of an exogenous variable on the dependent variable of all other units, which corresponds to the average column effect.

4. Bayesian estimation

In this section we present a Bayesian approach to estimating the parameters of the model introduced in Section 3. LeSage & Pace [20] includes Bayesian estimation for (2), but there are substantial differences between our implementation and that of [20], which will be clarified below. In Subsection 4.2, we also introduce a Bayesian implementation for the MESS error model (3). As already pointed out, our specification is slightly different from the one in [21], which requires $\mathbf{D}$ to be symmetric. Moreover, LeSage & Pace [21] focus their attention on ML estimation performing both a Monte Carlo simulation and an empirical illustration.

4.1. MESS model

In contrast to [20] (see also [19]), we do not use a Taylor series expansion with a fixed number of terms to approximate the matrix exponential, but we use instead the function 'expAtv' from the R package 'expm' (see [11]), which is an implementation of Sidje & Stewart [27] algorithm for the computation of the action of a matrix exponential on a vector[2]. The approach of the original MESS model implementation to approximating the matrix exponential was to use a Taylor polynomial of a fixed order p, which is the same for all computations of matrix exponentials involved in the estimation of the MESS model parameters. As a consequence, the truncation error in LeSage

[2] The R implementation is based on Fortran code by Sidje [26]

& Pace [20] varies by construction and might be either below or above a desired threshold level; conversely, the approach used here controls for the truncation error, which is guaranteed by construction to be below a chosen threshold. In Subsection 5.6 we study the efficiency of the method, comparing it to a number of other approximations to the matrix exponential.

In addition to using a different approach to the matrix exponential, we use different priors, choosing vague independence priors, rather than g-priors with a smaller and fixed variance, suggested in LeSage & Pace [20].

Prior and Posterior Distributions

As shown in [20], the log-likelihood of the model is equal to

$$l(\rho,\beta,\sigma^2;\mathbf{y},\mathbf{X}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \rho\,\mathrm{tr}(\mathbf{D}) - \frac{1}{2\sigma^2}(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta). \tag{10}$$

We choose the following prior distribution for σ^2:

$$\sigma^2 \sim \mathrm{InvGamma}(\kappa_0,\theta_0) \tag{11}$$

where κ_0 and θ_0 are small positive numbers. For the posterior distribution we obtain

$$\begin{aligned} \sigma^2 \mid \mathbf{y},\mathbf{X},\beta,\rho &\sim \mathrm{InvGamma}(\tilde{\kappa},\tilde{\theta}) \\ \tilde{\kappa} &= \kappa_0 + \frac{n}{2} \\ \tilde{\theta} &= \theta_0 + \frac{1}{2}(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta). \end{aligned} \tag{12}$$

We choose a diffuse prior for ρ: $p(\rho) \propto 1$.

Therefore the following holds for the conditional posterior

$$p(\rho|\mathbf{y},\mathbf{X},\beta,\sigma^2) \propto \exp\left[\rho\,\mathrm{tr}(\mathbf{D}) - \frac{1}{2\sigma^2}(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}}\mathbf{y}-\mathbf{X}\beta)\right]. \tag{13}$$

For the prior distribution of β we choose

$$\beta \sim \mathcal{N}(\beta_0, \tau^2\mathbf{H_0}), \quad \beta_0 = \mathbf{0}_k, \quad \mathbf{H_0} = 10^4 \cdot \mathbf{I}_k. \tag{14}$$

By equation (10):

$$\begin{aligned} \beta|\mathbf{y},\mathbf{X},\rho,\sigma^2 &\sim \mathcal{N}(\tilde{\beta},\tilde{\mathbf{H}}) \\ \tilde{\mathbf{H}} &= \left(\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}+\frac{\mathbf{H_0}^{-1}}{\tau^2}\right)^{-1} \\ \tilde{\beta} &= \tilde{\mathbf{H}}\left[\frac{\mathbf{X}'e^{\rho\mathbf{D}}\mathbf{y}}{\sigma^2}+\frac{\mathbf{H_0}^{-1}\beta_0}{\tau^2}\right]. \end{aligned} \tag{15}$$

The marginal distribution of ρ is obtained by integrating the following density:

$$p(\rho,\beta,\sigma^2|\mathbf{y},\mathbf{X}) \propto p(\mathbf{y}|\mathbf{X},\rho,\beta,\sigma^2)\cdot p(\rho,\beta,\sigma^2\mid\mathbf{X}) = p(\mathbf{y}|\mathbf{X},\rho,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)p(\rho), \tag{16}$$

with $p(\beta|\sigma^2)$ as in (14), $p(\sigma^2)$ as in (11). Following LeSage & Pace [20], we set $\tau^2=\sigma^2$ in (14).

This gives

$$p(\rho|\mathbf{y},\mathbf{X}) \propto p(\rho)\cdot\exp(\rho\,\mathrm{tr}(\mathbf{D}))\cdot(2\theta_0+H(\rho)+Q_1(\rho)+Q_2(\rho))^{-(\frac{n}{2}+\kappa_0)}, \tag{17}$$

where

$$\begin{aligned} &H(\rho)=(\mathbf{H}\exp(\rho\mathbf{D})\mathbf{y})'(\mathbf{H}\exp(\rho\mathbf{D})\mathbf{y}) && \mathbf{H}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &Q_1(\rho)=(\beta_0-\tilde{\beta})'\mathbf{H}_0^{-1}(\beta_0-\tilde{\beta}) && Q_2(\rho)=(\hat{\beta}-\tilde{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta}-\tilde{\beta}) \\ &\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\exp(\rho\mathbf{D})\mathbf{y} && \tilde{\beta}=\sigma^{-2}\tilde{\mathbf{H}}[\mathbf{X}'e^{\rho\mathbf{D}}\mathbf{y}+\mathbf{H_0}^{-1}\beta_0] \\ &\tilde{\mathbf{H}}=\sigma^2(\mathbf{X}'\mathbf{X}+\mathbf{H_0}^{-1})^{-1}. \end{aligned}$$

The R and C++ code developed for this paper apply an adaptive Metropolis Hastings algorithm, which is based on [29] [3], to draw samples from the marginal posterior density of ρ. Using this set of samples, we then obtain a posterior estimate of σ^2 and β by means of Gibbs sampling.

Durbin model

Passing to the spatially lagged regressor model (5), turns out to be equivalent to a reparametrization of model (2) as:

$$\exp(\rho\mathbf{D})\mathbf{y}=\mathbf{Z}\gamma+\varepsilon \tag{18}$$

[3] Available in an R package ([3])

where $\mathbf{Z} = [\mathbf{X}, \mathbf{DX}]$ and $\gamma' = (\beta', \theta')$.

Equations (10)–(17) then remain valid, once $\mathbf{X}$ is replaced by $\mathbf{Z}$ and β by the $2K-$dimensional vector γ (or $2K-1$ if the model includes the intercept and $\mathbf{D}$ is row-standardized contiguity matrix). Thus, for example, the posterior distribution of γ, writes:

$$\begin{aligned} \gamma|\mathbf{y}, \mathbf{X}, \rho, \sigma^2 &\sim \mathcal{N}(\tilde{\gamma}, \tilde{\mathbf{H}}) \\ \tilde{\mathbf{H}} &= \left(\frac{\mathbf{Z}'\mathbf{Z}}{\sigma^2} + \frac{\mathbf{H_0}^{-1}}{\tau^2} \right)^{-1} \\ \tilde{\gamma} &= \tilde{\mathbf{H}} \left[\frac{\mathbf{Z}' e^{\rho \mathbf{D}} \mathbf{y}}{\sigma^2} + \frac{\mathbf{H_0}^{-1} \gamma_0}{\tau^2} \right] \end{aligned} \tag{19}$$

where γ_0 and $\mathbf{H}_0$ are the hyperparameters of the prior distribution:

$$\gamma \sim \mathcal{N}(\gamma_0, \tau^2 \mathbf{H_0}), \quad \gamma_0 = \mathbf{0}_{2k}, \quad \mathbf{H_0} = 10^4 \cdot \mathbf{I}_{2k}$$

4.2. *MESS error model*

In this subsection we derive Bayesian estimation methods for the parameters of the MESS error model. As noted in Sections 3.2 and 4.2, by using a suitable reparametrization of the models, both the MESS error model with spatial splines (8) and the MESS Durbin error model (7) can be written as (6). For this reason, we limit ourselves to presenting the prior and posterior distributions for model (6), without loss of generality.

We provide a software implementation using the marginal posterior of ρ and an adaptive Metropolis Hastings algorithm (Spiegelhalter et al. [29], and Chivers [3]).

Conditionally on the parameters of model (6), $\mathbf{y}$ has the following distribution:

$$\mathbf{y}|\mathbf{X}, \beta, \rho, \sigma^2 \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 e^{-\rho \mathbf{D}} e^{-\rho \mathbf{D}'}). \tag{20}$$

Therefore, the likelihood function is equal to

$$p(\mathbf{y} \mid \mathbf{X}, \beta, \rho, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ \rho \operatorname{tr}(\mathbf{D}) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' (e^{\rho \mathbf{D}'} e^{\rho \mathbf{D}}) (\mathbf{y} - \mathbf{X}\beta) \right\}. \tag{21}$$

Prior and Posterior Distributions

We use vague priors for the computations, and independence priors for β. We choose the following prior distribution for σ^2:

$$\sigma^2 \sim \text{InvGamma}(\kappa_0, \theta_0) \tag{22}$$

where κ_0 and θ_0 are small positive numbers. For the posterior distribution we obtain

$$\begin{aligned} p(\sigma^2|\mathbf{y},\mathbf{X},\beta,\rho) &\propto p(\sigma^2)p(\mathbf{y}|\mathbf{X},\beta,\rho,\sigma^2) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n/2+\kappa_0+1} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}})(\mathbf{y}-\mathbf{X}\beta)]\right\} \exp\left(-\frac{\theta_0}{\sigma^2}\right) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\kappa_0+\frac{n}{2}+1} \exp\left\{-\tilde{\theta}/\sigma^2\right\} \end{aligned} \tag{23}$$

with $\tilde{\theta} = \theta_0 + \frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}})(\mathbf{y}-\mathbf{X}\beta)$. That is,

$$\sigma^2 \mid \mathbf{y},\mathbf{X},\beta,\rho \sim \text{InvGamma}(\tilde{\kappa},\tilde{\theta}), \;\; \tilde{\kappa} = \kappa_0 + \frac{n}{2}. \tag{24}$$

We choose the diffuse prior for ρ: $p(\rho) \propto 1$. Therefore, the following holds for the posterior

$$p(\rho|\mathbf{y},\beta,\sigma^2) \propto \exp\left[\rho \text{tr}(\mathbf{D}) - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\beta)'(e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}})(\mathbf{y}-\mathbf{X}\beta)\right]. \tag{25}$$

For the prior distribution of $\beta|\sigma^2$ we choose

$$\beta \sim \mathscr{N}(\beta_0, \tau^2\mathbf{H_0}), \quad \beta_0 = 0, \quad \mathbf{H_0} = 10^4 \cdot \mathbf{I}_n. \tag{26}$$

By equation (21), $\beta|\sigma^2,\rho,\mathbf{y},\mathbf{X} \sim \mathscr{N}(\bar{\beta},\bar{\mathbf{H}})$, with

$$\begin{aligned} \bar{\mathbf{H}} &= \left(\frac{\mathbf{X}'e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}}\mathbf{X}}{\sigma^2} + \frac{\mathbf{H_0}^{-1}}{\tau^2}\right)^{-1} \\ \bar{\beta} &= \bar{\mathbf{H}}\left[\frac{\mathbf{X}'e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}}\mathbf{y}}{\sigma^2} + \frac{\mathbf{H_0}^{-1}\beta_0}{\tau^2}\right]. \end{aligned}$$

We obtain the marginal posterior distribution of ρ by integrating the density

$$p(\rho,\beta,\sigma^2|\mathbf{y},\mathbf{X}) \propto p(\mathbf{y}|\mathbf{X},\rho,\beta,\sigma^2) \cdot p(\rho,\beta,\sigma^2 \mid \mathbf{X}) = p(\mathbf{y}|\mathbf{X},\rho,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)p(\rho) \tag{27}$$

with $p(\beta|\sigma^2)$ as in (26), $p(\sigma^2)$ as in (22). As in Subsection , we set $\tau^2 = \sigma^2$ in (26).

First, we rewrite the likelihood function:

$$\begin{aligned} p(\mathbf{y}|\mathbf{X},\rho,\beta,\sigma^2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\rho\,\mathrm{tr}(\mathbf{D}) - \frac{1}{2\sigma^2}(\mathrm{e}^{\rho\mathbf{D}}\mathbf{y} - \mathrm{e}^{\rho\mathbf{D}}\mathbf{X}\beta)'(\mathrm{e}^{\rho\mathbf{D}}\mathbf{y} - \mathrm{e}^{\rho\mathbf{D}}\mathbf{X}\beta)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\rho\,\mathrm{tr}(\mathbf{D}) - \frac{1}{2\sigma^2}\left(H(\rho) + (\beta-\hat{\beta})'\mathbf{X}'e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}}\mathbf{X}(\beta-\hat{\beta})\right)\right\}, \end{aligned} \tag{28}$$

with

$$\begin{aligned} H(\rho) &= (\mathbf{H}\exp(\rho\mathbf{D})\mathbf{y})'(\mathbf{H}\exp(\rho\mathbf{D})\mathbf{y}) & \mathbf{H} &= \mathbf{I} - \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}' \\ \mathbf{M} &= \exp(\rho\mathbf{D})\mathbf{X} & \hat{\beta} &= (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\exp(\rho\mathbf{D})\mathbf{y}. \end{aligned} \tag{29}$$

Therefore, because of $p(\rho) \propto 1$,

$$\begin{aligned} & p(\rho,\beta,\sigma^2|\mathbf{y},\mathbf{X}) \\ & \propto \frac{1}{\sigma^{n+k+2\kappa_0+2}} \cdot \exp\left\{\rho\,\mathrm{tr}(\mathbf{D}) - \frac{2\theta_0 + H(\rho) + Q_1(\rho) + Q_2(\rho) + (\beta-\tilde{\beta})'\bar{\mathbf{H}}^{-1}(\beta-\bar{\beta})}{2\sigma^2}\right\}, \end{aligned} \tag{30}$$

where

$$\bar{\beta} = \sigma^{-2}\bar{\mathbf{H}}[\mathbf{H}_0^{-1}\beta_0 + \mathbf{X}'e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}}\mathbf{X}\hat{\beta}] \tag{31}$$

$$\begin{aligned} \bar{\mathbf{H}} &= \sigma^2(\mathbf{X}'e^{\rho\mathbf{D}'}e^{\rho\mathbf{D}}\mathbf{X} + \mathbf{H_0}^{-1})^{-1} \\ Q_1(\rho) &= (\beta_0 - \bar{\beta})'\mathbf{H}_0^{-1}(\beta_0 - \bar{\beta}) \\ Q_2(\rho) &= (\hat{\beta} - \bar{\beta})'\mathbf{X}'\exp(\rho\mathbf{D}')\exp(\rho\mathbf{D})\mathbf{X}(\hat{\beta} - \bar{\beta}). \end{aligned} \tag{32}$$

Consequently, for the marginal posterior distribution of ρ, we have:

$$p(\rho\mid\mathbf{y}) = \int_0^\infty\int_{R^k}(\sigma^{-2})^{\tilde{\kappa}+1}\exp\left\{\rho\,\mathrm{tr}(\mathbf{D}) - \frac{2\theta_0 + H(\rho) + C(\rho) + (\beta-\bar{\beta})'\bar{\mathbf{H}}^{-1}(\beta-\bar{\beta})}{2\sigma^2}\right\}d\beta d\sigma^2$$

where $\tilde{\kappa} = (n+k)/2 + \kappa_0$,

$$C(\rho) = Q_1(\rho) + Q_2(\rho) = \hat{\beta}'\mathbf{X}'\mathrm{e}^{\rho\mathbf{D}'}\mathrm{e}^{\rho\mathbf{D}}\mathbf{X}\hat{\beta} + \beta_0'\mathbf{H}_0^{-1}\beta_0 - \bar{\beta}'\bar{\mathbf{H}}^{-1}\bar{\beta}.$$

Then, by integrating with respect to β

$$\begin{aligned} p(\rho \mid \mathbf{y}) &= e^{\rho \mathrm{tr}(\mathbf{D})} \int_0^\infty (\sigma^{-2})^{\frac{n+k+2\kappa_0+2}{2}} \exp\left\{ -\frac{2\theta_0 + H(\rho) + C(\rho)}{2\sigma^2} \right\} \left(\sqrt{2\pi\sigma^2}\right)^k |\bar{\mathbf{H}}|^{1/2} d\sigma^2 \\ &\propto e^{\rho \mathrm{tr}(\mathbf{D})} |\bar{\mathbf{H}}|^{1/2} \int_0^\infty (\sigma^{-2})^{\frac{n+2\kappa_0+2}{2}} \exp\left\{ -\frac{\tilde{\theta}(\rho)}{\sigma^2} \right\} \\ &\propto e^{\rho \mathrm{tr}(\mathbf{D})} |\bar{\mathbf{H}}|^{1/2} (\tilde{\theta}(\rho))^{-(n/2+\kappa_0)} \end{aligned} \tag{33}$$

where $\tilde{\theta}(\rho) = (2\theta_0 + H(\rho) + Q_1(\rho) + Q_2(\rho))/2$.

In most spatial applications, the weight matrices are assumed to have zero diagonal elements, therefore the factor $\exp(\mathrm{trace}(\mathbf{D}))$ normally disappears. However, the formulas in this section allow for different choices of the weight matrix and thus our implementations extends to more general contexts.

We have implemented the model using the programming language R, together with code written in C++ and included with the help of the R-packages 'Rcpp' and 'Rcpp - Armadillo' ([8, 9]). As in the MESS model (2), we use an adaptive Metropolis Hastings algorithm ([29]) available in the R-package 'MHadaptive' ([3]), to sample from the marginal posterior density (33).

5. Application

This section describes the application of the different MESS models to house price data from Scotland[4].

Considering aggregate data for each region, we apply the MESS model to an analysis of the dependence of log house prices (in 1000 GBP) on the median number of rooms, the log crime rate (log of number of recorded crimes per 10,000 living in the area), sales (ratio of sales of houses to total number of houses), the logarithm of the average time (in minutes) it takes to reach the nearest shopping centre by car, and the house type predominant in the area ("detached", "semi", "flat", "terrace"). The use of these independent variables is suggested by Lee [15].

[4]The data were originally taken from the Scottish Neighbourhood Statistics ([25]) database (http://www.sns.gov.uk/) and converted to spatial data frames for R to be used with the 'CARBayes' package ([16])

The region we study is that of the Glasgow and Clyde health board, the data are from the year 2008. The region is divided into 270 intermediate geographies (IG), which are "small areas that have a median area of 124 hectares and a median population of 4,239" (see Lee [15]).

5.1. MESS model

First we estimate the parameters of the MESS model with the row-standardized contiguity matrix of the given lattice as the weight matrix. We draw 50,000 samples from the marginal posterior distribution of ρ, with a burn-in of 5000. We use a thinning factor of 10. The Gibbs sampler for β and σ^2 assumes a burn-in of 4,500. We assume the parameter τ, in (14) equal to σ. Table 2 summarizes the result. We compute the DIC using p_D as in Spiegelhalter et al. [30], (originally introduced in [29]) and p_V [30], where

$$\begin{aligned} \overline{D(\theta)} &= \mathbb{E}_{\theta|\mathbf{y}}[-2\log(p(\mathbf{y}|\theta))] \qquad & p_D &= \overline{D(\theta)} + 2\log[p(\mathbf{y}|\mathbb{E}(\theta|\mathbf{y}))] \\ D(\theta|y) &= -2\log(p(\mathbf{y}|\theta)) \qquad & p_V &= \frac{1}{2}\text{var}(D(\theta|\mathbf{y})) \end{aligned} \quad (34)$$

p_V was also suggested by Gelman et al. [10].

Following [32], for observation i, CPO_i is defined as the density of $i-th$ observation, y_i, given all the other observations, and it is estimated as the harmonic mean of the density $f(y_i|\rho,\beta,\sigma,\rho)$ computed at each iteration. As intuitively clear, larger CPO_i values are preferred, and as an overall CPO we consider the product of CPO_i.

As a benchmark, the second column in Table 2 reports the estimates of the Bayesian linear regression (that is, with $\rho = 0$). The linear regression parameters are estimated using the same priors for β and σ^2 as we used for the estimation of the MESS parameters. [5]

[5] Very similar results would be obtained by taking ML MESS estimates as benchmark, for example running the package "spdep" by Eric Blankmeyer based on the SE Toolbox spatial/mess.m implementation. The extension of the comparison to ML implementation and estimation is however beyond the scope of this work.

	Row-standardized contiguity		Linear regression		7-nearest neighbors	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
ρ	-0.290	[-0.414,-0.164]			-0.523	[-0.682,-0.368]
β_0	3.697	[2.950,4.469]	5.196	[4.655,5.719]	2.989	[2.283,3.703]
β_1	0.188	[0.133, 0.243]	0.197	[0.140,0.254]	0.179	[0.126,0.231]
β_2	-0.121	[-0.187,-0.056]	-0.166	[-0.230,-0.100]	-0.122	[-0.182,-0.061]
β_3	0.0021	[0.0014 0.0028]	0.0021	[0.0014, 0.0028]	0.0021	[0.0015, 0.0028]
β_4	-0.071	[-0.120,-0.023]	-0.090	[-0.139, -0.038]	-0.041	[-0.088,0.006
β_5	-0.120	[-0.227, -0.012]	-0.156	[-0.273,-0.049]	-0.138	[-0.239, -0.036]
β_6	-0.190	[-0.307,-0.071]	-0.227	[-0.349,-0.106]	-0.186	[-0.297,-0.074]
β_7	-0.259	[-0.388, -0.128]	-0.280	[-0.414, -0.144]	-0.268	[-0.393,-0.145]
σ^2	0.0466	[0.0392,0.0553]	0.050	[0.042,0.060]	0.042	[0.035,0.050]

Table 2: Estimation with MESS model and Bayesian linear regression, β_0 = intercept, β_1 = number of rooms, β_2 = log crime rate, β_3 = sales, β_4 = log time to shop, β_5 = type "semi", β_6 = type "flat", β_7 = type "terrace". Mean values and 95% credible intervals.

The corresponding DIC is -29.2 with $p_D = 8.8$, as compared to a DIC of -53.2 with $p_D = 9.6$ for the MESS model. Thus according to the DIC criterion, MESS is preferable to Bayesian linear regression.

We compute the DIC of the MESS model with different weight matrices. For k-nearest neighbors weight matrices, the lowest DIC is obtained with $k = 7$. We also compute the DIC of the model with the binary instead of the row-standardized contiguity matrix, and obtain a DIC equal to -81.5 with $p_D = 9.4$.

Unlike the DIC, the analysis with CPOs indicates the row-standardized contiguity matrix to be preferable to the 7-nearest neighbors weight matrix. We obtain a value, on the log scale, of 77.97 for the MESS model with row-standardized contiguity, while the Bayesian linear model on the log scale produces a CPO equal to 25.03, and the MESS model with 7-nearest neighbors matrix a CPO equal to 9.75.

In Table 3 direct, indirect and total effects are shown. The results reported show that the estimates of the direct and indirect effects of the MESS model, as well as

	Row-standardized contiguity		
	Direct	Indirect	total
number of rooms	0.189 [0.133, 0.246]	0.062 [0.031, 0.1]	0.251 [0.174, 0.335]
log crime rate	-0.122 [-0.188, -0.055]	-0.039 [-0.067, -0.017]	-0.161 [-0.245, -0.075]
sales	0.002 [0.001, 0.003]	0.001 [0 ,0.001]	0.003 [0.002, 0.004]
log time to shop	-0.072 [-0.121, -0.025]	-0.023 [-0.044, -0.007]	-0.095 [-0.159, -0.033]
type ”semi”	-0.122 [-0.23, -0.014]	-0.039 [-0.084 ,-0.004]	-0.162 [-0.304, -0.019]
type ”flat”	-0.193 [-0.313 ,-0.071]	-0.063 [-0.119 ,-0.021]	-0.256 [-0.419, -0.095]
type ”terrace”	-0.263 [-0.4 ,-0.135]	-0.086 [-0.157, -0.035]	-0.349 [-0.537, -0.179]
	7-nearest neighbors		
	Direct	Indirect	total
number of rooms	0.182 [0.126, 0.233]	0.119 [0.07, 0.182]	0.3 [0.206, 0.404]
log crime rate	-0.124[-0.185, -0.061]	-0.081 [-0.134, -0.039]	-0.205 [-0.31, -0.103]
sales	0.002 [0.002, 0.003]	0.001 [0.001, 0.002]	0.004 [0.002, 0.005]
log time to shop	-0.042 [-0.09, 0.007]	-0.027 [-0.059, 0.005]	-0.069 [-0.147, 0.012]
type ”semi”	-0.14 [-0.241, -0.036]	-0.092 [-0.176, -0.022]	-0.232[-0.406 ,-0.058]
type ”flat”	-0.189 [-0.305 ,-0.075]	-0.124 [-0.22 ,-0.048]	-0.313 [-0.515, -0.124]
type ”terrace”	-0.273 [-0.399, -0.146]	-0.179 [-0.3, -0.083]	-0.452 [-0.68, -0.239]

Table 3: Direct, indirect and total effects for the MESS (lag) model with Row-standardized contiguity and 7-nearest neighbors matrices. 95% credible intervals in parentheses.

the estimates of the parameters, are quite robust to the choice of the weight matrices. However, it must be kept in mind that differences between specifications could have been mitigated by a low spatial autocorrelation.

In order to quantify the impact of the issue raised by Rodrigues et al. [24], namely the occurrence of opposite sign of marginal and partial correlations in the MESS model, we computed the proportion of negative partial correlations between regions (the marginal correlations are in fact all positive). For the model with the row-standardized contiguity matrix, this proportion, computed in the posterior estimation, is 0.248. That is, about 25% of all the partial correlations are negative. However, as the spatial MESS correlation in this application is relatively low, the negative partial correlations are also close to irrelevant. In particular, the proportion of negative correlations smaller than -0.01 among all negative correlations is about 0.02 (for the posterior mean of ρ).

For the model with the 7-nearest neighbors weight matrix, the proportion of negative partial correlations between regions is quasi identical to that of the model with the row-standardized contiguity matrix, being 0.247. However, the negative correlations are higher, in absolute value; in fact the proportion of negative correlations lower than -0.01 among all negative correlations is 0.095 (for the posterior mean of ρ). A similar computation, where fictitious regions are defined on a regular lattice and a row-standardized adjacency matrix is used, is presented in the Appendix, for both the MESS and the SAR covariances.

5.1.1. MESS Durbin model

Here we are applying the Mess Durbin model as specified by equation (5). For the spatial weight matrices we use both the row-standardized contiguity matrix and the 7-nearest neighbors matrix, and results are shown in Table 4.

The estimates are in most cases comparable to those of the corresponding models without lagged regressors. The only exception is the estimate of the parameter and of the direct and indirect effects of the average time to reach a shopping centre: it is positive and not significant, for all the weight matrices used, whereas the same quantities estimated from the MESS model without lagged regressors was negative and significant. Both the models estimated show some evidence of spatially lagged effects; in particular, the estimates in Table 4 suggest the log house price to be affected by the distance from a shopping center mainly at neighboring locations.

Observing the DIC (using *pD*) results, -87.87 and -94.57 respectively for the row-standardized contiguity matrix and the 7-nearest neighbors matrix, it looks like the introduction of the spatially lagged regressors improve the model. However, the CPO results do not indicate this improvement, while, again, row-standardized contiguity matrix performs better than the 7-nearest neighbors matrix, with a CPO (in log scale) respectively of 20.19 and 18.42.

5.2. MESS error model

5.2.1. MESS Durbin error model

Here we are applying the Mess Durbin error model as specified by equation (7). For the spatial weight matrix we use both the row-standardized contiguity matrix and the 7-nearest neighbors matrix, results are shown in Table 6. The estimated parameters are in line with the estimates from the MESS Durbin model, with a few discrepancies. For example, with the row-standardized matrix, the lagged effect of the number of rooms is negative and significant in the MESS Durbin model, while it is not significant (but still negative) in the MESS Durbin error model.

5.3. MESS error model with spatial splines

Here we apply the MESS error model with spatial splines introduced in Section 3.2. For the spatial weight matrix we use a row-standardized contiguity matrix, in order to avoid problems with the interpretation of the parameter k of the k-nearest neighbors matrices in an irregular lattice (see i.e. [28])

In this particular example the coefficients of the components of the tensor product of the splines are all fairly symmetrically distributed around a mean close to zero. Therefore, we repeat the estimation including only the individual splines and not their tensor product, thereby reducing the DIC and increasing predictive ability (see Table 7). We also implement the model with splines of the individual coordinates with five degrees of freedom, again without including the tensor product of the splines. The

	Row-standardized contiguity		7-nearest neighbors	
	Mean	95%CI	Mean	95% CI
ρ	-0.48	[-0.643, -0.32]	-0.631	[-0.857 ,-0.427]
β_0	4.388	[3.381, 5.448]	3.37	[1.82, 4.952]
β_1	0.206	[0.153 ,0.259]	0.191	[0.138, 0.242]
β_2	-0.12	[-0.182, -0.056]	-0.123	[-0.185, -0.062]
β_3	0.002	[0.001, 0.003]	0.002	[0.002, 0.003]
β_4	0.012	[-0.045, 0.071]	0.013	[-0.046, 0.07]
β_5	-0.169	[-0.269, -0.065]	-0.145	[-0.245, -0.044]
β_6	-0.26	[-0.375 ,-0.141]	-0.217	[-0.332, -0.102]
β_7	-0.313	[-0.437, -0.193]	-0.295	[-0.419, -0.169]
θ_1	-0.149	[-0.255, -0.049]	-0.083	[-0.213, 0.045]
θ_2	-0.13	[-0.225, -0.035]	-0.057	[-0.218 ,0.111]
θ_3	-0.001	[-0.003, 0]	-0.001	[-0.003 , 0]
θ_4	-0.145	[-0.236 ,-0.056]	-0.124	[-0.217 ,-0.03]
θ_5	0.04	[-0.156, 0.229]	-0.053	[-0.324, 0.215]
θ_6	0.184	[-0.016, 0.382]	0.116	[-0.142, 0.374]
θ_7	0.223	[-0.035, 0.474]	0.14	[-0.202 ,0.479]
σ^2	0.04	[0.034, 0.048]	0.04	[0.033, 0.047]

Table 4: Estimation with MESS (Lag) Durbin model, β_0 = intercept, β_1 = number of rooms, β_2 = log crime rate, β_3 = sales, β_4 = log time to shop, β_5 = type ”semi”, β_6 = type ”flat”, β_7 = type ”terrace”. Mean values and 95% credible intervals.

	Row-standardized contiguity		
	Direct	Indirect	total
number of rooms	0.211 [0.156, 0.266]	0.123 [0.07, 0.193]	0.335 [0.237, 0.447]
log crime rate	-0.123 [-0.187, -0.057]	-0.071 [-0.122, -0.031]	-0.194 [-0.301, -0.09]
sales	0.002 [0.001, 0.003]	0.001 [0.001, 0.002]	0.003 [0.002, 0.005]
log time to shop	0.012 [-0.046, 0.072]	0.007 [-0.027, 0.046]	0.02 [-0.074, 0.116]
type ”semi”	-0.173 [-0.275, -0.067]	-0.101 [-0.187, -0.035]	-0.273 [-0.451, -0.105]
type ”flat”	-0.266 [-0.385, -0.144]	-0.156 [-0.268, -0.069]	-0.421 [-0.636, -0.22]
type ”terrace”	-0.321 [-0.448, -0.197]	-0.188 [-0.315, -0.092]	-0.509 [-0.74,9 -0.3]
	7-nearest neighbors		
	Direct	Indirect	total
number of rooms	0.196 [0.142, 0.249]	0.165 [0.088, 0.274]	0.361 [0.23 0.523]
log crime rate	-0.127 [-0.19, -0.064]	-0.106 [-0.19, -0.046]	-0.233 [-0.38 -0.11]
sales	0.002 [0.002, 0.003]	0.002 [0.001, 0.003]	0.004 [0.003 0.006]
log time to shop	0.013 [-0.047, 0.071]	0.011 [-0.04, 0.067]	0.024 [-0.087 0.138]
type ”semi”	-0.149 [-0.252, -0.045]	-0.125 [-0.248, -0.034]	-0.274 [-0.5 -0.079]
type ”flat”	-0.223 [-0.342, -0.104]	-0.188 [-0.341, -0.074]	-0.411 [-0.683 -0.178]
type ”terrace”	-0.303[-0.432, -0.173]	-0.256 [-0.447, -0.118]	-0.559 [-0.879 -0.291]

Table 5: Direct, indirect and total effects for the MESS (Lag) Durbin model with Row-standardized contiguity and 7-nearest neighbors matrices. 95% credible intervals in parentheses.

	Row-standardized contiguity		7-nearest neighbors	
	Mean	95% CI	Mean	95% CI
ρ	-0.584	[-0.772, -0.402]	-0.611	[-0.726, -0.496]
β_0	6.587	[5.323, 7.849]	5.313	[3.227, 7.327]
β_1	0.196	[0.142, 0.251]	0.196	[0.142, 0.249]
β_2	-0.143	[-0.207, -0.079]	-0.127	[-0.188, -0.062]
β_3	0.002	[0.001, 0.003]	0.002	[0.002, 0.003]
β_4	-0.008	[-0.063, 0.046]	-0.005	[-0.06, 0.05]
β_5	-0.169	[-0.276, -0.063]	-0.144	[-0.247, -0.042]
β_6	-0.248	[-0.365, -0.129]	-0.205	[-0.32, -0.09]
β_7	-0.307	[-0.437 ,-0.175]	-0.288	[-0.413, -0.16]
θ_1	-0.072	[-0.187, 0.044]	0.079	[-0.096, 0.26]
θ_2	-0.206	[-0.319, -0.095]	-0.111	[-0.333, 0.106]
θ_3	0	[-0.002, 0.001]	0	[-0.002, 0.002]
θ_4	-0.167	[-0.272, -0.057]	-0.119	[-0.252, 0.022]
θ_5	0.007	[-0.223 ,0.232]	-0.109	[-0.459, 0.231]
θ_6	0.118	[-0.109, 0.347]	0.059	[-0.292, 0.403]
θ_7	0.062	[-0.224, 0.35]	-0.025	[-0.448, 0.403]
σ^2	0.04	[0.033, 0.048]	0.047	[0.041, 0.054]

Table 6: Estimation with MESS Durbin error model, β_0 = intercept, β_1 = number of rooms, β_2 = log crime rate, β_3 = sales, β_4 = log time to shop, β_5 = type "semi", β_6 = type "flat", β_7 = type "terrace". Mean values and 95% credible intervals.

	3 nodes with tensor product		3 nodes		5 nodes	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
ρ	-0.498	[-0.689,-0.315]	-0.552	[-0.734,-0.378]	-0.437	[-0.63, -0.259]
β_0	4.336	[-0.947, 9.775]	4.592	[4.1,5.07]	4.481	[3.847, 5.119]
β_1	0.221	[0.171, 0.273]	0.222	[0.17, 0.274]	0.221	[0.171, 0.273]
β_2	-0.076	[-0.139, -0.013]	-0.082	[-0.145, -0.018]	-0.093	[-0.153, -0.034]
β_3	0.0024	[0.002, 0.003]	0.0022	[0.0016, 0.0029]	0.0023	[0.0017, 0.0029
β_4	0.029	[-0.03, 0.088]	-0.027	[-0.086, 0.03]	0.012	[-0.046, 0.069]
β_5	-0.167	[-0.267, -0.065]	-0.168	[-0.265, -0.063]	-0.165	[-0.265, -0.069]
β_6	-0.281	[-0.396, -0.164]	-0.279	[-0.396, -0.16]	-0.243	[-0.353, -0.132]
β_7	-0.309	[-0.429, -0.187]	-0.327	[-0.450, -0.199]	-0.29	[-0.41,-0.173]
σ^2	0.039	[0.033,0.047]	0.044	[0.037,0.052]	0.037	[0.032, 0.045]

Table 7: MESS error model with spatial splines. Mean values and 95% credible intervals.

results are included in Table 7. Finally, we compare this last model to the following model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{X}^*\beta^* + \varepsilon \tag{35}$$

where $\mathbf{X}^*$ is the basis matrix of natural splines of the coordinates of the centroids of the regions of the lattice. Like with the spatial error MESS model with splines above, we use natural splines of the individual coordinates with 5 degrees of freedom, but do not include the tensor product in the regression. The DICs with p_D and p_V and CPOs are listed in Table 8, together with those of the models discussed above. Ideally the number of degrees of freedom and location of the nodes of the splines of the coordinates should be estimated from the model. Biller [2] and DiMatteo et al. [7] suggested methods in this direction, but for the MESS model, such an approach would be considerably more intensive computationally. In line with the results presented in Subsection 5.1, the proportion of negative partial correlation corresponding to the MESS error model with five nodes for the splines is about 25% on average.

Despite the fact that the MESS error model does not allow for spatial spillover

Model	DIC using pD	pD	DIC using pV	pV	CPO
MESS err. mod., 3 nodes, tensor prod.	-82.4	23.97	-75.55	30.84	18.39
MESS err. mod., 3 nodes	-77.9	16.90	-77.47	17.26	16.57
MESS err. mod., 5 nodes	-99.62	19.9	-93.2	27.25	32.26
No MESS component, 5 nodes	-92.6	18.8	-91.0	20.4	31.54

Table 8: Spatial Splines with and without MESS component: DIC and CPO (in log scale)

interactions between the dependent variable nor spatial spillover effects, the MESS error model with splines outperforms the MESS (lag) specifications in terms of both the DIC and the CPO. This supports our claim that spatial splines are a flexible and promising way to capture different forms of heterogeneous effects in the model.

5.4. MESS and SAR

LeSage & Pace [20] and LeSage & Pace [21] introduce the MESS model in particular as an alternative to the spatial autoregressive SAR model. We will show that in practical applications the SAR and MESS models often lead to similar parameter estimates. However, it should be noted that the MESS and the SAR model have different correlation patterns by construction. The larger the spatial parameter, the greater are these differences [6]. In addition, we would like to point out the following: because of the complicated structure of the spatial parameter within the matrix exponential, a precise interpretation of the spatial parameter is impossible. The spatial parameter has a more straightforward interpretation in the SAR model.

We implement the SAR model with identical priors to compare the parameter estimates to those obtained with the MESS model. The parameter estimates are listed in Table 9, while direct, indirect and total effects are shown in Table 10. We compare the DIC and p_D of the two approaches, getting a DIC equal to -55.0 with $p_D = 9.9$ for the SAR model with the row-standardized contiguity matrix, and a DIC of -83.8 with

[6] Differences and similarities between correlation structures and effects induced by MESS and SAR models are further explored in the Appendix

	Row-standardized contiguity		7-nearest neighbors	
	Mean	95% CI	Mean	95% CI
ρ	0.290	[0.179, 0.398]	0.461	[0.345,0,571]
β_0	3.483	[2.670, 4.301]	2.695	[1.938, 3.483]
β_1	0.187	[0.132, 0.241]	0.179	[0.126, 0.230]
β_2	-0.117	[-0.183, -0.052]	-0.117	[-0.175, -0.058]
β_3	0.0021	[0.0014, 0.0028]	0.0021	[0.0015, 0.0028]
β_4	-0.067	[-0.117, -0.0189]	-0.037	[-0.085, 0.0087]
β_5	-0.116	[-0.223, -0.006]	-0.136	[-0.238, -0.034]
β_6	-0.183	[-0.302, -0.066]	-0.181	[-0.293, -0.0732]
β_7	-0.256	[-0.383, -0.128]	-0.262	[-0.384, -0.140]
σ^2	0.0454	[0.0382, 0.0541]	0.0403	[0.0339,0.0478]

Table 9: SAR model estimates. Mean values and 95% credible intervals.

$p_D = 9.9$ for the SAR model with the 7-nearest neighbors matrix. Thus, in our application, the SAR model is comparable to the MESS model in predictive power, although CPOs are considerably lower, with a CPO (in log scale) of 8.58 for the SAR with the row-standardized contiguity matrix and CPO of -1.94 for the 7-nearest neighbor.

The proportion of negative partial correlations between regions, computed in the posterior estimation, is 0.259, among which the proportion of correlations smaller than -0.01 is about 0.003 (for the posterior mean of ρ). These computations suggest the claim, supported by Theorem 1 and by the simple example in the Appendix, that, relative to the MESS model, the SAR model tends to induce a possibly larger number of negative partial correlations, although they tend to be lower in absolute value.

5.5. Simulation

Starting from a simple lattice consisting of 100 squares in a square, we simulate observation from a MESS model and a SAR model, and estimate a MESS model, a MESS error model with spatial splines, and SAR model. The MESS model and MESS error model are able to estimate the correct value of ρ when data are generated from a

	Row-standardized contiguity		
	Direct	Indirect	total
number of rooms	0.19 [0.135, 0.245]	0.074 [0.037, 0.123]	0.264 [0.183, 0.352]
log crime rate	-0.119 [-0.185, -0.053]	-0.046 [-0.081, -0.02]	-0.165 [-0.254, -0.076]
sales	0.002 [0.001, 0.003]	0.001 [0, 0.001]	0.003 [0.002, 0.004]
log time to shop	-0.069 [-0.118, -0.019]	-0.027 [-0.053, -0.007]	-0.096 [-0.166, -0.027]
type ”semi”	-0.118 [-0.228, -0.006]	-0.045 [-0.098, -0.003]	-0.163 [-0.315, -0.009]
type ”flat”	-0.187 [-0.308, -0.067]	-0.072 [-0.137, -0.023]	-0.259 [-0.43, -0.096]
type ”terrace”	-0.261 [-0.39, -0.13]	-0.101 [-0.187, -0.041]	-0.362 [-0.555, -0.178]
	7-nearest neighbors		
	Direct	Indirect	total
number of rooms	0.185 [0.13, 0.238]	0.146 [0.103, 0.188]	0.331 [0.233, 0.426]
log crime rate	-0.121 [-0.181, -0.06]	-0.096 [-0.143, -0.048]	-0.217 [-0.324, -0.108]
sales	0.002 [0.002, 0.003]	0.002 [0.001, 0.002]	0.004 [0.003, 0.005]
log time to shop	-0.038 [-0.088, 0.009]	-0.03 [-0.069, 0.007]	-0.068 [-0.157, 0.016]
type ”semi”	-0.141 [-0.246, -0.035]	-0.111 -[0.195, -0.028]	-0.252 [-0.441, -0.063]
type ”flat”	-0.188 [-0.303, -0.076]	-0.148 [-0.24, -0.06]	-0.336 [-0.542, -0.136]
type ”terrace”	-0.271 [-0.397, -0.144]	-0.214 [-0.314, -0.114]	-0.485 [-0.711, -0.258]

Table 10: Direct, indirect and total effects for the SAR model with Row-standardized contiguity and 7-nearest neighbors matrices. 95% credible intervals in parentheses.

MESS model.

Moreover, the promising result is that both the MESS model and MESS error model with spatial splines turn out to be flexible and to estimate the parameter ρ quite coherently with the value of the autocorrelation parameter λ when data are generated from a SAR model, although, as expected, they perform worse than SAR itself. The higher the value λ of the SAR model, the worse is the estimation. For example, when 100 data set are generated from a SAR model with $\lambda = 0.8$, fitting a MESS error model with spatial splines, the average estimate of ρ is -1.019 ($95\%CI = [-1.269, -0.786]$), with average DIC of 141. According to the formula suggested by [20], this should roughly correspond to $\rho \approx -1.609$. However, this conversion formula is purely indicative and presents some serious limitations, some of which are pointed out in the Appendix. Fitting a MESS model, we obtain an average estimate for ρ of -1.06 ($95\%CI = [-1.347, -0.799]$), with an average DIC of 123. Finally, fitting a SAR model, we obtain an average estimate for λ of 0.662 ($95\%CI = [0.633, 0.795]$), with average DIC of 73. Direct and indirect effects estimated from the different models are always comparable, suggesting a flexibility of the MESS model to grasp the correct spatial mechanism.

5.6. Computational Efficiency

In Table 11 the different approaches to the computation of the matrix exponential are compared in terms of efficiency. The 'expm' package ([11]) includes a number of different algorithms to compute the matrix exponential. The default method implemented in the package is 'Higham08.b'. This is an implementation of Higham's algorithm (see [13], algorithm 10.20), with an extra balancing step developed by Stadelmann [31]. We use the package 'rbenchmark' ([14]) for a comparison of computing times of different algorithms on the same laptop computer.

We have included in the comparison not only the functions from the 'expm' package, but also the function 'expmat' from the Armadillo C++ library, which we access through 'RccpArmadillo' ([9]). In the list below, this function is referred to as myExpoMat. We use the row-standardized weights matrix $\mathbf{D}$ to compare computing times for different approximations to the matrix exponential $\exp(0.5\mathbf{D})$ implemented in the

method	replications	relative time	elapsed time
expAtv(D, log.house.price, 1)	10	1.000	0.355
expm(D, method = ”Higham08.b”)	10	3.006	1.067
myExpoMat(D)	10	3.513	1.247
expm(D, method = ”PadeRBS”)	10	3.538	1.256
expm(D, method = ”Higham08”)	10	3.569	1.267
expm(D, method = ”Taylor”, order = 2)	10	4.789	1.700
expm(D, method = ”Taylor”, order = 3)	10	5.101	1.811
expm(D, method = ”Taylor”, order = 4)	10	5.403	1.918
expm(D, method = ”Taylor”, order = 6)	10	5.696	2.022

Table 11: Comparison of the efficiency of different approaches to the computation of the matrix exponential

R package ’expm’ ([11]). **D** is the adjacency matrix of the example above. The matrix **D** in the list below is equal to 0.5**D**).

The method ’expAtv’ from the ’expm’ package stands out as the fastest algorithm. This algorithm does not compute explicitly the exponential of a matrix, but directly the action of this exponential on a vector, in this case a vector of logarithms of house prices.

’Higham08.b’ is the fastest of those methods which directly compute the matrix exponential, and not its action on a vector.

It should be noted that with the not extremely sparse matrices of an intermediate dimension of 270×270 in our application, there was no gain in computational speed by using the sparse matrix format. Calling the package ’Matrix’ our program can run using a sparse matrix representation.

6. Conclusion

The matrix exponential model with one parameter was introduced as an alternative to the widely used SAR model. This paper addresses a number of concerns related to the MESS model, thereby evaluating its performance through an analysis based on theory, and on predictive accuracy in an application. We analyze the model both on its own, and in relation to its natural competitor, the SAR model. Moreover, we develop a new implementation of Bayesian parameter estimation for the MESS model with vague prior distributions, which is shown to be precise and computationally efficient. We try to offer a complete schematized view of the possible specifications based on the use of the matrix exponentials, thus covering the classes of the so called lag models (the standard MESS model belongs to this class) and the error models. Focusing in particular on error models (although in principle the same idea may be applied to lag models) we propose a generalization, based on the introduction of spatial splines, that is able to capture spatial heterogeneities not accounted for in the MESS error framework.

By computing the estimated marginal and partial correlations from the MESS model corresponding to different choices of the weight matrix, we find that the fraction of pairs of locations with opposite signs is generally just below 25%, which is almost identical of the fraction of opposite signs in the estimated SAR correlation matrices. This finding is coherent with Theorem 1 in the Appendix, and mitigates the severity of this particular problematic aspect of the MESS model, at least in the context of comparing the MESS to the SAR model.

In terms of parameter estimation, the MESS model is comparable to the SAR approach, and estimates are quite robust to the choices of the weight matrix. The predictive accuracies of the MESS and SAR models are close, thus showing that, in general, there is no clear superiority of one approach with respect to the other. At least in this particular case, neither the SAR nor the MESS patterns are able to capture the propagation on the endogenous variable, which appears to depend, in a complex way, on a direct spatial effect of the observable and on exposure to unobserved exogenous shocks. The model including spatial splines among the regressors outperforms both the MESS and the SAR in terms of predictive accuracy. Similar encouraging results are found by performing several simulations, where the model with splines proves to be more robust to model misspecification. This suggests the approach based to the introduction of spatial splines to be a flexible way to cope with model uncertainty.

R code to fit the MESS and MESS error models (with/without splines) using the approach described in this paper is available.

References

[1] Anselin, L., & Florax, R. (1995). *New Directions in Spatial Econometrics*. Springer.

[2] Biller, C. (2012). Adaptive Bayesian regression splines in semiparametric generalized linear models. *J Comput Graph Stat*, *9*, 122 – 140.

[3] Chivers, C. (2012). *MHadaptive: General Markov Chain Monte Carlo for Bayesian inference using adaptive Metropolis-Hastings sampling*. URL: `http://CRAN.R-project.org/package=MHadaptive` r package version 1.1-8.

[4] Ciu, T. Y. M., Leonard, T., & Tsui, K. (1996). The matrix-logarithmic covariance model. *J Am Stat Assoc*, *91*, 198 – 210.

[5] Debarsy, N., Jin, F., & fei Lee, L. (2015). Large sample properties of the matrix exponential spatial specification with an application to {FDI}. *Journal of Econometrics*, *188*, 1 – 21. doi:`http://dx.doi.org/10.1016/j.jeconom.2015.02.046`.

[6] Deng, X., & Tsui, K. (2013). Penalized covariance matrix estimation using a matrix-logarithmic transformation. *J Comput Graph Stat*, *22*, 494 – 512.

[7] DiMatteo, I., Genovese, C., & Kass, R. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, *88*, 1055 – 1071.

[8] Eddelbuettel, D., Francois, R., Allaire, J. J., Ushey, K., Bates, D., & Chambers, J. (2015). *Rcpp: seamless R and C++ Integration*. URL: `http://CRAN.R-project.org/package=Rcpp` r package version 0.11.5.

[9] Francois, R., Eddelbuettel, D., & Bates, D. (2015). *RcppArmadillo: Rcpp integration for the Armadillo templated linear algebra library*. URL: `http://CRAN.R-project.org/package=RcppArmadillo` r package version 0.4.650.1.1.

[10] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2003). *Bayesian data analysis*. (2nd ed.). Chapman & Hall/CRC.

[11] Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., Stadelmann, M., & expm-developers@lists.R-forge.R-project.org (2014). *expm: Matrix exponential*. URL: `http://CRAN.R-project.org/package=expm` r package version 0.99-1.1.

[12] Han, X., & fei Lee, L. (2013). Model selection using j-test for the spatial autoregressive model vs. the matrix exponential spatial model. *Reg. Sci. and Urban. Eco*, *43*, 250–271.

[13] Higham, N. J. (2008). *Functions of Matrices: Theory and Computation*. SIAM.

[14] Kusnierczyk, W. (2012). *rbenchmark: Benchmarking routine for R*. URL: `http://CRAN.R-project.org/package=rbenchmark` r package version 1.0.0.

[15] Lee, D. (2013). CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Stat. Softw.*, *55*, 1–24. URL: `http://www.jstatsoft.org/v55/i13/`.

[16] Lee, D. (2014). *CARBayes: spatial areal unit modelling*. URL: `http://CRAN.R-project.org/package=CARBayes` r package version 4.0.

[17] Leonard, T., & Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *Ann Stat*, *20*, 1669 – 1696.

[18] Leorato, S., & Mezzetti, M. (2016). Spatial panel data model with error dependence: A bayesian separable covariance approach. *Bayesian Analysis*, *11*, 1035–1069.

[19] LeSage, J. P. (2015). Spatial econometrics toolbox. URL: `http://www.spatial-econometrics.com/`.

[20] LeSage, J. P., & Pace, R. K. (2007). A matrix exponential spatial specification. *J Econometrics*, *140*, 190 – 214.

[21] LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

[22] Pavlides, M. G., & Perlman, M. D. (2009). How likely is Simpson's paradox? *Am Stat*, *63*, 226 – 233.

[23] Piribauer, P., & Fischer, M. M. (2015). Model uncertainty in matrix exponential spatial growth regression models. *Geogr Anal*, *47*, 240–261.

[24] Rodrigues, E., Assunçãu, R., & Dey, D. K. (2014). A closer look at the spatial exponential matrix specification. *Spat. Statist.*, *9*, 109–21.

[25] Scottish neighbourhood statistics (Accessed 17 February 2015). URL: `http://www.sns.gov.uk/`.

[26] Sidje, R. B. (1998). Expokit. a software package for computing matrix exponentials. *ACM T Math Software*, *24*, 130 – 156.

[27] Sidje, R. B., & Stewart, W. J. (1999). A numerical study of large sparse matrix exponentials arising in Markov chains. *Comput Stat Data An*, *29*, 345 – 68.

[28] Smirnov, O., & Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Comput Stat Data An*, *35*, 301 – 319.

[29] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J Roy Stat Soc B*, *64*, 583 – 639.

[30] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *J Roy Stat Soc B*, *76*, 485 – 493.

[31] Stadelmann, M. (2009). Matrixfunktionen: Analyse und Implementierung [matrix functions: analysis and implementation, in German]. URL: `ftp://ftp.sam.math.ethz.ch/pub/sam-reports/reports/reports2009/2009-12.pdf`.

[32] Stern, H. S., & Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, *19*, 2377–2397.

[33] Wagner, C. H. (1982). Simpson's paradox in real life. *Am. Stat.*, *36*, 46 – 48.

[34] Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*, 434–449.

A. Comparing MESS and SAR models

One of the potentially problematic aspects of the MESS model is the lack of interpretability of the parameter ρ. Another concern recently raised by [24] is related to a somewhat unexpected behavior of the MESS covariance matrix, which implies in rather frequent cases, opposite sign of partial and marginal correlations.

In this section we briefly discuss these problems, emphasizing differences and similarities between MESS and SAR models. We focus, in particular, on the role of the parameter ρ compared to the autoregressive parameter, and on how this affects the pattern of the covariances as well as the marginal effects.

In order to avoid confusion concerning notation, we use the letter λ for the spatial correlation parameter in a SAR model.

The role of ρ and the marginal effects

The difficult interpretation of the parameter ρ of the MESS model has been stated by LeSage & Pace [20] themselves, who proposed to consider the norms of the matrixes $S(\rho) = \exp(\rho\mathbf{D})$ and $S^*(\lambda) = \mathbf{I} - \lambda\mathbf{D}$. For a row-standardized contiguity matrix $\mathbf{D}$, the norms $\|\exp(\rho\mathbf{D})\|_\infty = \mathrm{e}^\rho$, and $\|\mathbf{I} - \lambda\mathbf{D}\|_\infty = 1 - \lambda$ suggest that the parameter ρ corresponding to a given spatial autoregressive parameter could be read as a monotone transformation of the autocorrelation coefficient, i.e. $e^\rho \approx 1 - \lambda$.

This approximated relation is very useful to facilitate interpretation of ρ. However, it is not helpful in identifying differences in the marginal effects of lag models since they are unit-specific and change with the pattern of the covariance matrices.

In fact, if we denote by $a(i,j)$ and $a^*(i,j)$ the $(i,j)-$th elements of $S(\log(1-\lambda))^{-1} = \exp\{-\log(1-\lambda)\mathbf{D}\}$ and $S^*(\lambda)^{-1}$ respectively, then, $i = 1,\ldots,n$

$$\sum_{j=1}^{n} a(i,j) = \sum_{j=1}^{n} a^*(i,j).$$

Since, for the SAR models, the direct and indirect effects follow from Table 1, once we replace a by a^*, the above identity implies that the total impact of variable k is the same for the two models: $\bar{M}_{tot}(k) = \bar{M}^*_{tot}(k)$ for all k, where $M^*_{tot}(k) = \sum_{i,j} a^*(i,j)\beta_h$; however, in general, $\bar{M}_{dir}(k) \neq \bar{M}^*_{dir}(k)$ and $\bar{M}_{in}(k) \neq \bar{M}^*_{in}(k)$.

To get an idea of how the two specifications determine different effects, consider the matrix $\mathbf{D}$ corresponding to a time series: then the $(i,j)-$th component of $\mathbf{D}^h$ – $d_h(i,j)$ – is equal to one if $j = i+h$ and is zero otherwise. Then, the MESS and SAR effects, $\partial y_i/\partial x_{jk}$, are equal to: $\beta_k a(i,j) = \beta_k \frac{(-\log(1-\lambda))^{i-j}}{(i-j)!}$ and $\beta_k a^*(i,j) = \beta_k \lambda^{i-j}$ respectively, for $i > j$. Therefore, the MESS model has unbounded indirect effects in a neighborhood of 1, while the corresponding absolute effect of the SAR models is always bounded above by $|\beta_k|$. While the absolute effects are symmetric in λ for the SAR model, for the MESS model, if $\lambda < 0$ and $\rho = \log(1-\lambda)$, $|\beta_j a(i,l)| < |\beta_j|(\log(2)) < |\beta_j|$.

Simpson's paradox and the MESS and SAR models

The MESS model is centrally based on a very specific spatial correlation structure induced by the matrix exponential. The necessity of analyzing the correlation pattern was also emphasized by Rodrigues et al. [24], who noticed a striking peculiarity concerning the MESS correlation pattern, namely, the fact that negative partial correlations tend to occur frequently in the framework of the MESS model. Note that, since the errors are assumed to be normally distributed, the partial correlations may be defined as conditional correlations. The correlation structure of the MESS model depends, of course, on the specific weight matrix used. Apart from contiguity, other spatial weight matrices, in particular k-nearest neighbors, have been used with the MESS model.

Opposite signs of marginal and partial correlation also occur in this case, provided that the weight matrix can be represented by a simple graph (which excludes nonzero diagonal elements). Let $\mathbf{D} = \{d(i,j)\}_{ij}$ be the weight matrix used for a particular MESS model and let us write $\mathbf{D}^h = \{d_h(i,j)\}_{ij}$. This weight matrix may be a contiguity weight matrix, a k-nearest neighbors matrix, or a different type of weight matrix, which is usually chosen to be sparse.

Theorem 1. *Let $\mathbf{D}$ be weight matrix with symmetric zero entries, that is $d(i,j) = 0$ if and only if $d(j,i) = 0$, and consider the covariance matrix $\Sigma_\nu = [(I-\lambda\mathbf{D})^{-1}(I-\lambda\mathbf{D}')^{-1}]$ of a SAR model. Then for two regions i and j marginal and partial correlations have different signs, if $d_{2k+1}(i,j) = 0$ for all k.*

The above result is analogous to that proved by Rodrigues et al. [24] for the MESS model. We point out that, differently from Rodrigues et al. [24], the weight matrix **D** is not required to be symmetric. Similar aguments apply to the MESS covariance matrix, thus extending the result of Rodrigues et al. [24] to many types of non-symmetric weight matrices. The condition of symmetry of zero entries is in particular satisfied by the row-standardized version of any symmetric distance matrix. The fact that under the same assumptions of Theorem 1 both the SAR and the MESS marginal and partial correlations have opposite signs, attenuates the findings of Rodrigues et al. [24], although limited to the particular cases considered here (that is regular lattices with particular weight matrices).

Proof. Since $d_h(i,j) = \sum_{m=1}^{n} d_l(i,m)d_{h-l}(m,j) = 0$ if and only if, either $d_l(i,m) = 0$ or $d_{h-l}(m,j) = 0$, for all m and for any $l \leq h$, then we can conclude that, for all l,k such that $l+k$ is odd, $\sum_{m=1}^{n} d_l(i,m)d_k(j,m) = 0$, exploiting the fact that $d_k(j,m) = 0$ if and only if $d_k(m,j) = 0$.

Then, under the assumptions of Theorem, the $(i,j)-$th element of the covariance matrix Σ_ν writes (for $i \neq j$):

$$\begin{aligned} [\Sigma_\nu]_{ij} &= \sigma^2 \left[(I-\lambda\mathbf{D})^{-1}(I-\lambda\mathbf{D}')^{-1}\right]_{ij} = \sigma^2 \left[\sum_{h=0}^{\infty}\lambda^h\mathbf{D}^h\sum_{k=0}^{\infty}\lambda^k\mathbf{D}'^k\right]_{ij} \\ &= \sigma^2\left[\sum_{k=0}^{\infty}\sum_{l=0}^{2k}\lambda^{2k}\sum_{m=1}^{n} d_l(i,m)d_{2k-l}(j,m)\right]_{ij} > 0 \end{aligned} \tag{36}$$

and

$$[\Sigma_\nu^{-1}]_{ij} = \frac{1}{\sigma^2}\lambda^2\sum_m d_1(i,m)d_1(j,m) > 0$$

As ν is Gaussian[7], the sign of the partial correlation between ν_i and ν_j is opposite to that of $[(\Sigma_\nu)^{-1}]_{ij}$. Therefore, the partial correlation between ν_i and ν_j is negative, while the marginal one is positive. It then follows that, also in the case of the SAR specification, the partial correlation is negative, while the marginal correlation is positive.

[7]Recall that, for a Gaussian random vector $Y_1,\ldots,Y_n$ with precision matrix Ω, the (i,j)th partial correlation (for $i \neq j$) is equal to $-\omega(i,j)/\sqrt{\omega(i,i)\omega(j,j)}$.

Of course, Theorem 1 provides a sufficient, but not a necessary condition for the partial and marginal correlations of both the SAR and MESS models to have opposite sign. In fact, negative partial correlation (conditional on all the other outcomes) is likely to occur whenever two regions are not odd order neighbors up to a finite order k_0. Moreover, while each term of the sums defining Σ_v and Σ_v^{-1} takes the form of an infinite series of powers of $\mathbf{D}$, the precision matrix of the SAR is a finite sum, thus the partial (conditional) correlation of (i, j) is zero whenever $d_l(i, j) = 0$ for $l \leq 2$, while the marginal correlation is, in general, nonzero.

The particular values of ρ and λ for which these considerations do not fully apply, depend, of course, on the structure of the lattice and the weight matrix used.

In general, a low absolute value of ρ or of λ is not necessarily able to prevent negative partial correlations or to reduce the number of their occurrences. However, it also leads to low values of these correlations. Therefore, the relevance of negative partial correlations in practical applications of the MESS model depends on the strength of the spatial correlations in rather complicated ways. It should be noted that opposite signs of partial and marginal correlations do occur in practice, and are referred to as Simpson's paradox (see [33] and [22]). Therefore, their existence itself is nothing to worry about.

A simple example

As the spatial parameter ρ is inside the matrix exponential, the correlation pattern is highly complicated and somewhat non-intuitive. Therefore, in this section we illustrate the correlation pattern of the MESS model, in comparison with the SAR model, on a very simple lattice consisting of 16 squares in a square.

We consider the binary contiguity matrix of the lattice and its corresponding adjacency row normalized weight matrix as in Figure 1.

We compute the correlation matrix of the MESS and SAR models associated with this lattice for $\lambda = 0.8$ and $\lambda = 0.5$ in Figure 2 and in Figure 3. In both cases, $\rho = \log(1-\lambda)$.

From this example we learn about a number of interesting characteristics of the MESS model. As noted in the previous section, the contribution of powers of the pa-

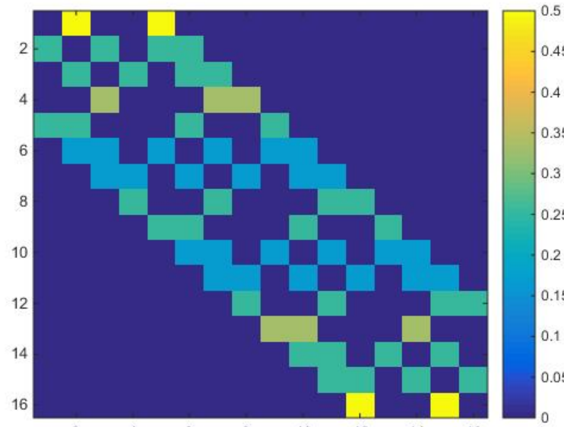

Figure 1: Adjacency Row Normalized Weight Matrix

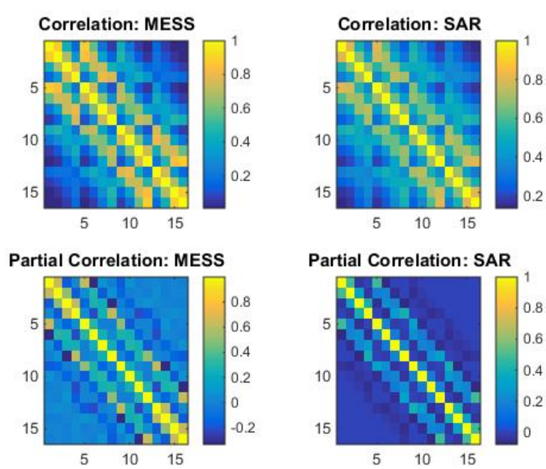

Figure 2: Marginal and Partial correlation, $\lambda = 0.8$, $\rho = \log(1-\lambda)$

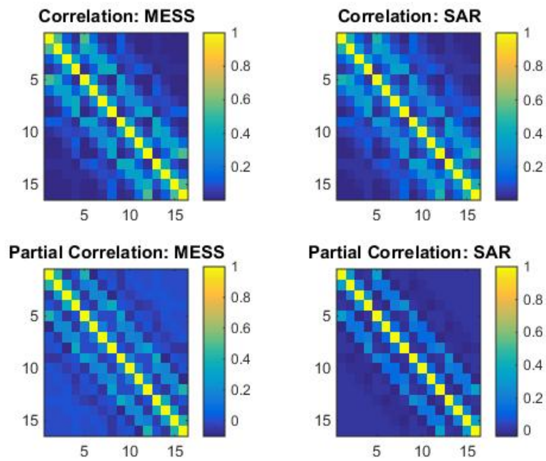

Figure 3: Marginal and Partial correlation with $\lambda = 0.5$, $\rho = \log(1-\lambda)$

rameter ρ in the covariance decays exponentially, as opposed to the geometric decay characterizing the autoregressive SAR model. This causes the spatial correlation between two regions to decrease faster with distance: for example, assuming $\rho = -\log 5$ the correlation between regions 1 and 2 is around 0.88, while that between regions 1 and 3 is around 0.44, compared to the corresponding values of 0.81 and 0.52 of the SAR model with λ equal 0.8. On the other hand, assuming $\rho = -\log 2$ the correlation between regions 1 and 2 is around 0.53, while that between regions 1 and 3 is around 0.11, compared to the corresponding values of 0.46 and 0.131 of the SAR model with λ equal 0.5.

Both the MESS and the SAR models assume negative partial correlations, for pairs of regions (i, j) at odd distances. In particular, assuming $\rho = -\log 5$, the fraction of negative partial correlation for the MESS correlation model is 39.84%, while for the SAR model, assuming $\lambda = 0.8$, the fraction of negative partial correlation for the MESS correlation model is 51.56%. These fractions increase to 43.75% and 54.69%, for the MESS and SAR correlations respectively, when $\lambda = 0.5$ (and $\rho = -\log 2$). However, MESS correlations are much higher in absolute values, with a 22.66% being below -0.05, while for the SAR only the 6.25% lies below that threshold and only for

$\lambda = 0.8$.

B. Computation of impact measures

Even when $\mathbf{D}$ is a sparse matrix, the exponential $\exp\{\mathbf{D}\}$ in general is not, and the computation of the total and direct marginal effects requires the computation of a dense $n \times n$ matrix. However, just as in the SAR effects case, if $\mathbf{D}$ is a row-standardized matrix, the computation of the total effects is dramatically simplified by the fact that $\iota^\top \mathbf{D}^j \iota = n$ for all $j \geq 0$ and thus the total effect of the kth covariate for the MESS models are[8]:

MESS	$n^{-1}\iota^\top \exp\{-\rho\mathbf{D}\}\iota\beta_k = n^{-1}\sum_j \frac{\rho^j}{j!}\iota^\top\iota\beta_k = e^{-\rho}\beta_k$
MESS Durbin	$n^{-1}\left(\iota^\top \exp\{-\rho\mathbf{D}\}\iota\beta_k + \iota^\top\mathbf{D}\exp\{-\rho\mathbf{D}\}\iota\theta_k\right) = e^{-\rho}(\beta_k + \theta_k)$

So, in case of a row-standardized matrix, the trace of $\exp\{-\rho\mathbf{D}\}$ is the main computational problem one has to face. We recall that, if the matrix $\mathbf{D}$ is diagonalizable, $\mathrm{trace}(\exp\{-\rho\mathbf{D}\}) = \sum_j e^{-\rho\lambda_j}$, where λ_j are the eigenvalues of $\mathbf{D}$. The computation of the eigenvalues of $\mathbf{D}$ is manageable, and once it is done, the λ_j need not be recomputed within the MCMC iterations.

In particular, for the MESS model, the direct effect of the kth covariate is equal to $n^{-1}\sum_j e^{-\rho\lambda_j}\beta_k$, while, for the MESS Durbin model, it is

$$\begin{aligned}
n^{-1}\sum_j e^{-\rho\lambda_j}\beta_k + n^{-1}\sum_{h=1}^{\infty}\frac{\rho^h}{h!}\mathrm{trace}(\mathbf{D}^{h+1})\theta_k &= n^{-1}\sum_j e^{-\rho\lambda_j}\beta_k + n^{-1}\sum_{h=1}^{\infty}\frac{\rho^h}{h!}\sum_{j=0}^{n}\lambda_j^{h+1}\theta_k \\
&= n^{-1}\sum_j e^{-\rho\lambda_j}\beta_k + n^{-1}\sum_{j=1}^{n}\lambda_j e^{-\rho\lambda_j}\theta_k
\end{aligned}$$

[8]The effects of MESS error models are straightforward