Approximate Bayesian Inference for Spatial Econometrics Models

Roger S. Bivand[a], Virgilio Gómez-Rubio[b,*], Håvard Rue[c]

[a] *Department of Economics, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway.*
[b] *Department of Mathematics, School of Industrial Engineering, University of Castilla-La Mancha, 02071 Albacete, Spain.*
[c] *Department of Mathematical Sciences, Norwegian University for Science and Technology, N-7491 Trondheim, Norway.*

Abstract

In this paper we explore the use of the Integrated Laplace Approximation (INLA) for Bayesian inference in some widely used models in Spatial Econometrics. Bayesian inference often relies on computationally intensive simulation methods, such as Markov Chain Monte Carlo. When only marginal inference is needed, INLA provides a fast and accurate estimate of the posterior marginals of the parameters in the model.

Furthermore, we have compared the results provided by these models to those obtained with a more general class of Generalised Linear Models with random effects. In these models, spatial autocorrelation is modelled by means of correlated Gaussian random effects.

We also discuss a procedure to extend the class of models that the **R-INLA** software can fit. This approach is based on conditioning on one or more parameters so that the resulting models can be fitted with **R-INLA** across sets of values of the fixed parameters. The posterior marginals of these parameters of interest are then obtained by combining the marginal likelihoods (which are conditioned on the values of the parameters fixed) of the fitted models and a prior on these parameters. This approach can also be used to fit even more general models.

Finally, we discuss the use of all these models on two datasets based on median housing prices for census tracts in Boston and the probability of business re-opening in New Orleans in the aftermath of hurricane Katrina.

Keywords: Bayesian inference, INLA, Spatial Econometrics

*Corresponding author

Email addresses: roger.bivand@nhh.no (Roger S. Bivand), Virgilio.Gomez@uclm.es (Virgilio Gómez-Rubio), hrue@math.ntnu.no (Håvard Rue)

1. Introduction

Economic data often shows spatial patterns, for example housing prices are similar across adjacent neighbourhoods, or GDP varies smoothly across regions in countries. Spatial econometrics models aim at including this spatial dependence so that the value of an observation depends on the observed values of its neighbours.

Traditionally, for a continuously observed variable, such as housing prices, this dependence has been expressed explicitly, i.e., the price at an observed location is centred at a weighted average of the observed values at its neighbours plus perhaps the effect of other covariates at the observation. Several autoregressive models described by Cliff and Ord [4] have been built on this idea, as described in Section 2.1. When the response is non-Gaussian, autoregressive models are difficult to handle, as discussed in Section 2.2. In particular, we will consider the case of a binary outcome with a Bernoulli distribution, but this approach can easily be extended to other variables from the Exponential family.

Bayesian hierarchical models provide a slightly different approach to the same problem by considering a spatially-structured latent random effect to account for spatial correlation. The random effects specification accounts for spatial autocorrelation in the disturbances rather than among the observed responses.

In all cases, the resulting models can be complex and model fitting becomes a problem. In a Bayesian framework, inference is often based on computationally intensive methods such as Markov Chain Monte Carlo (MCMC) to obtain the joint posterior distribution of the parameters, see, for example, Besag et al. [3]. Once this has been obtained, it is easy to compute summary statistics of the model parameters, credible intervals and other quantities of interest.

When only marginal inference is needed, other methods are available. Rue et al. [23] describe the Integrated Nested Laplace Approximation (INLA) to obtain an approximation to the posterior marginals of the parameters of interest. They also provide the **R-INLA** software (`http://www.r-inla.org`) to fit a wide range of models which in most cases reduce computation time to seconds and allow for the use of larger datasets. See Eidsvik et al. [5] for a discussion of the benefits of using INLA on large datasets.

Our aim is to apply this new methodology to spatial econometrics models. LeSage and Pace [13] provide a full description of the most important models in Spatial Econometrics, which they fit using a Bayesian approach. Furthermore, they provide software for Bayesian inference with MCMC in their Spatial Econometrics Toolbox (`http://www.spatial-econometrics.com/`). In this work, we also consider Bayesian inference but using a completely different model fitting technique based on INLA.

Although many of the spatial econometrics models cannot be fitted with **R-INLA**, we have devised a new approach for fitting models (after conditioning on some parameters) using the **R-INLA** software. These models are then combined using Bayesian Model Averaging to obtain the posterior marginals of the required model. It should be stressed that, even though we focus on spatial econometrics models, the approach presented in this paper can be used to fit many other models not explicitly included in **R-INLA**.

The paper is organised as follows. In Section 2 we provide a summary of the spatial econometrics models used in this paper. Section 3 describes Integrated Nested Laplace Approximation for approximate Bayesian inference. Some computational details needed to fit spatial econometrics models with **R-INLA** are given in Section 4. A simulation study comparing model fitting under different assumptions has been included in Section 5. Two examples are discussed in Section 6. In Section 7 we have included a discussion on other importants models and issues on Spatial Econometrics. Finally, Section 8 includes a general discussion of the paper.

2. Spatial Econometrics Models

Spatial models have been used in spatial econometrics for a long time, see Anselin [2] for a review. In general, the interest in spatial econometrics is on modelling spatial interaction in an autoregressive way, so that the observation at a given area, y_i, depends on a weighted sum of the values of the variable at its neighbours plus some other (fixed) effects and some random noise. This makes the spatial dependence explicit but it also introduces a particular variance-covariance structure in the error term of the models, as seen below.

2.1. Gaussian models

A popular model is the Simultaneous Autoregressive (SAR) model (or Spatial Error Model, SEM model), which can be expressed as follows:

$$y = X\beta + u; u = \rho_{\text{Err}} W u + e; e \sim MVN(0, \sigma^2 I_n). \tag{1}$$

Here, $y = (y_1, \ldots, y_n)$ is the vector of observations, X is a design matrix of p covariates, $\beta = (\beta_1, \ldots, \beta_p)$ are the coefficients of the covariates, I_n is the identity matrix of dimension $n \times n$ and W is a row-standardised adjacency matrix. ρ_{Err} is a parameter that measures spatial autocorrelation. Furthermore, the error terms are modelled to be a weighted sum of the random errors at their neighbours plus some random noise e, which is multivariate Normal (MVN) with zero mean and diagonal variance-covariance matrix $\sigma^2 I_n$. LeSage and Pace [13] also describe a variation of this model where the variance-covariance matrix of the error term is diagonal with n different variance parameters (i.e., $\sigma_1^2, \ldots, \sigma_n^2$). We will not discuss this class of models in this paper, but the resulting models can be easily derived from the equations included here.

The SEM model can be re-written as

$$y = X\beta + e'; e' \sim MVN(0, \sigma^2 (I_n - \rho_{\text{Err}} W)^{-1} (I_n - \rho_{\text{Err}} W')^{-1}). \tag{2}$$

In this case, the model is a general linear regression with a non-diagonal variance-covariance matrix for the error term.

Another important model in spatial econometrics is the Spatial Lag model (SLM model), which is referred to as the SAR model in LeSage and Pace [13].

In this model, the response is modelled as depending on a weighted sum of the responses at their neighbours, plus a linear term on covariates and an error term:

$$y = \rho_{\text{Lag}} W y + X\beta + e; e \sim MVN(0, \sigma^2 I_n). \quad (3)$$

This model can be rewritten as:

$$y = (I_n - \rho_{\text{Lag}} W)^{-1} X\beta + e';\ e' \sim MVN(0, \sigma^2 (I_n - \rho_{\text{Lag}} W)^{-1} (I_n - \rho_{\text{Lag}} W')^{-1}). \quad (4)$$

Finally, a third model that is widely used in spatial econometrics is the Spatial Durbin model (SDM model):

$$y = \rho_{\text{Lag}} W y + X\beta + WX\gamma + e = [X, WX][\beta, \gamma] + e;\ ; e \sim MVN(0, \sigma^2 I_n). \quad (5)$$

γ is a vector of coefficients for the spatially lagged covariates WX, so that the response now depends not only on the (weighted) values of the response at neighbours and the covariates, but also on the (weighted) values of the covariates at neighbours.

The Spatial Durbin model can be expressed as a Spatial Lag model as follows:

$$y = \rho_{\text{Lag}} W y + X^* \beta' + e;\ X^* = [X, WX];\ \beta' = [\beta, \gamma], \quad (6)$$

and

$$y = (I_n - \rho_{\text{Lag}} W)^{-1} X^* \beta' + e';\ e' \sim MVN(0, \sigma^2 (I_n - \rho_{\text{Lag}} W)^{-1} (I_n - \rho_{\text{Lag}} W')^{-1}). \quad (7)$$

In all these models, the error term is Gaussian with zero mean and variance-covariance matrix given by a SAR specification. The Spatial Lag and Durbin models have also a more complex structure in the linear term on the covariates which needs to be dealt with. Furthermore, the spatial autocorrelation parameter ρ (regardless of the model used) is restricted to be in the range $(1/\lambda_{min}, 1/\lambda_{max})$, where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of W, respectively. When W is row-standardised this implies that $\lambda_{max} = 1$ and $\rho < 1$. See Haining [10] for details.

LeSage and Pace [13] provide Matlab code to fit all these models and obtain estimates and summary statistics of the model parameters. Using MCMC can sometimes be computationally intensive and simulations may take a while.

While these models are latent Gaussian, they have not been implemented within the **R-INLA** software. In the next section we will discuss how to overcome this problem by fitting these models by repeatedly conditioning on the values of ρ. This is a general approach which could be used for other models that cannot be fitted with **R-INLA** but that become "R-INLA fittable" once some parameters (usually, one or two) have been fixed.

2.2. Non-Gaussian models

When the response variable is binary (i.e., the outcome is zero or one), the previous models are not adequate. LeSage et al. [14] provide a Spatial Probit model to estimate the probability of re-opening a business in New Orleans after hurricane Katrina. They model the outcome y_i as follows:

$$y_i = \begin{cases} 1 & if\ y_i^* \geq 0 \\ 0 & if\ y_i^* < 0 \end{cases}, \tag{8}$$

where y_i^* is a latent variable which measures the latent net profit (i.e., if it is higher than zero the business will re-open). This latent variable is modelled in turn using a Spatial Lag model:

$$y^* = \rho_{\text{Lag}} W y^* + X\beta + e;\ e \sim MVN(0, \sigma^2 I_n). \tag{9}$$

We will use the following expression to represent this model

$$y^* = (I_n - \rho_{\text{Lag}} W)^{-1} X\beta + e';\ e' \sim N(0, \sigma^2 (I_n - \rho_{\text{Lag}} W)^{-1} (I_n - \rho_{\text{Lag}} W')^{-1}), \tag{10}$$

as it can be used with the Generalised Linear Models described below. Also, it is clear from equation (8) that the relationship between y_i and y_i^* is non-linear.

2.3. Generalised Linear Models with random effects

A different way of modelling the outcome and accounting for covariates and spatial autocorrelation is by means of the Generalised Linear Models [17]. The outcome y_i is assumed to come from a distribution of the Exponential family with mean parameter μ_i. The relationship between μ_i and the linear predictor on a vector of covariates X_i is established through a link function $g(\cdot)$:

$$g(\mu_i) = \eta_i = X_i\beta. \tag{11}$$

Spatial dependence is included in the model by means of correlated random effects:

$$\eta_i = X_i\beta + u_i, \tag{12}$$

where $u = (u_i, \ldots, u_n)$ is multivariate Normal distributed with zero mean and variance-covariance matrix Σ. Different structures for Σ have been proposed to model spatial dependence. In previous models a SAR specification has been considered:

$$\Sigma = \sigma^2 (I_n - \rho W)^{-1} (I_n - \rho W')^{-1}. \tag{13}$$

Another widely used specification is the Conditionally autoregressive (CAR) specification:

$$\Sigma = \sigma^2 (I_n - \rho W)^{-1}. \tag{14}$$

Note that now W is a symmetric matrix, and not a row-standardised one, in order to have a valid variance-covariance matrix.

Model fitting and interpretation of this kind of models is somewhat easier in the SAR probit as the effects of the covariates and spatial correlation are included in different terms. In the Spatial Lag model, for example, the term on the covariates is

$$(I - \rho_{\text{Lag}} W)^{-1} X\beta, \tag{15}$$

clearly showing that the effect of a covariate depends on its coefficients β and the spatial correlation ρ_{Lag}.

Also, if a spatial lag on the covariates is required in the model (as in the Spatial Durbin model) it can be added as follows:

$$\eta_i = X_i\beta + WX_i\gamma + u_i. \tag{16}$$

The models described in Sections 2.1 and 2.2 can be expressed as Generalised Linear Models. The Gaussian models described in Section 2.1 are Gaussian GLMs with random effects after conditioning on ρ. The variance-covariance matrix is a SAR specification. Similarly, for a given value of ρ, the Spatial Probit model is a binomial GLM with a probit link function. Other models to deal with a binary response include spatial filtering with an autologistic model [9].

In other words, the Spatial Econometrics models described before can be fitted using methods and standard software for Generalised Linear Mixed-effects models conditioning on ρ. Based on this insight, we provide a way of fitting these models with **R-INLA** in Section 4.

Bayesian inference for these models often requires the use of Markov Chain Monte Carlo techniques. LeSage and Pace [13] provide Matlab code to fit these and many other models in their Spatial Econometrics Toolbox. Running this code for large datasets however can be time consuming and our aim is to find alternative ways of providing Bayesian inference. For this reason we will rely on the Integrated Nested Laplace Approximation, which is discussed in the next Section.

2.4. Impacts

LeSage and Pace [13] and LeSage et al. [14] discuss how changes of a covariate at location i will affect the output at location j. To measure these effects, they define *direct* and *indirect* impacts.

For the case of the Gaussian models, the impacts are defined as the partial derivatives of the linear predictor at site i, η_i on $x_{v,j}$, the value of covariate v at site j:

$$\frac{\partial \eta_i}{\partial x'_{v,j}}. \tag{17}$$

The matrix of impacts for covariate v can be easily derived but will depend on the model. For the SEM model, it is

$$\frac{\partial \eta}{\partial x_v'} = I_n \beta_v;\ v = 1, \ldots, p. \tag{18}$$

For the SLM model the impacts become

$$\frac{\partial \eta}{\partial x_v'} = (I_n - \rho_{\text{Lag}} W)^{-1} \beta_v;\ v = 1, \ldots, p\ , \tag{19}$$

while for the Spatial Durbin model we have

$$\frac{\partial \eta}{\partial x_v'} = (I_n - \rho_{\text{Lag}} W)^{-1} (\beta_v + W\gamma_v);\ v = 1, \ldots, p. \tag{20}$$

The average *total* impact is defined as the average (over n, the number of observations) of the $n \times n$ impacts. The average *direct* impact associated to covariate v is defined as the average of the diagonal of the previous matrix. This measures the average impact of changing covariate v at site i on the same site. The average *indirect* impact is defined as the average (over n) of the off-diagonal elements. Hence, this measures the average effect of changing covariate v at a site on any other site.

The case of the spatial probit is slightly different. Now the impacts are based on computing the partial derivatives of

$$\frac{\partial Pr(y_i = 1)}{\partial x_{v,j}'}, \tag{21}$$

which involves a non-linear term due to the fact that the response and the linear predictor are connected via a link function (i.e., the probit link). In this case the matrix of partial derivatives for the SEM model is

$$\frac{\partial Pr(y = 1)}{\partial x_v'} = D(f(\eta))\beta_v;\ v = 1, \ldots, p. \tag{22}$$

Here $f(\eta)$ is a vector of the standard Normal distribution evaluated at the values of the linear predictors $\eta_i, i = 1, \ldots, n$. $D(\cdot)$ simply represents a diagonal matrix made from its argument.

For the SLM we have

$$\frac{\partial Pr(y = 1)}{\partial x_v'} = D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1} \beta_v;\ v = 1, \ldots, p\ , \tag{23}$$

and for the SDM

$$\frac{\partial Pr(y = 1)}{\partial x_v'} = D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1} (\beta_v + W\gamma_v);\ v = 1, \ldots, p. \tag{24}$$

Direct and indirect effects are defined from the matrix of partial derivatives as in the Gaussian case. See LeSage and Pace [13], LeSage et al. [14] for details on how the impacts and effects are derived. They also provide some computational hints on how to compute the direct and indirect effects. In particular they avoid computing $(I_n - \rho W)^{-1}$ and instead they use an approximation of order k of the form

$$(I_n - \rho W)^{-1} \simeq \sum_{i=0}^{k} (\rho W)^i. \tag{25}$$

The posterior distribution of the impacts can be computed when the Bayesian model is fitted using MCMC. In this case, the impacts are computed at each iteration with the current values of the model parameters, so that at the end of the simulations we have a sample of the posterior distribution of the impacts.

LeSage and Pace [13, page 39] define the different average impacts as

$$\overline{M}(r)_{direct} = n^{-1} tr(S_r(W)) \tag{26}$$

$$\overline{M}(r)_{total} = n^{-1} i_n^{'} (S_r(W)) i_n \tag{27}$$

$$\overline{M}(r)_{indirect} = \overline{M}(r)_{total} - \overline{M}(r)_{direct}. \tag{28}$$

Here $S_r(W)$ is the matrix of partial derivatives with respect to x_r, $tr(\cdot)$ is the trace of a matrix and i_n is an n vector of 1's.

Table 1 summarises the different values of $S_r(W)$ and direct and indirect impacts for the spatial econometrics models discussed before. The traces that appear in two of the direct effects can be computed as

$$\begin{aligned} tr\Big((I_n - \rho W)^{-1}\Big) &\simeq tr(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^k W^k) = \\ &= tr(I_n) + \rho tr(W) + \ldots \rho^k tr(W^k) \end{aligned} \tag{29}$$

$$tr\Big((I_n - \rho W)^{-1} W\Big) \simeq tr(W) + \rho tr(W^2) + \ldots + \rho^k tr(W^{k+1}). \tag{30}$$

Hence, the traces of $W^i; i = 1, \ldots, k$ can be used at the same time to compute the direct effects for the SLM and SDM models.

The expressions of the total effects for the Probit models are similar to those of the Gaussian models but taking into account that $S_r(W)$ now includes the term $D(f(\eta))$. The total effects are those of the Gaussian case multiplied by $\overline{f(\eta)} = \sum_{i=1}^n \frac{f(\eta_i)}{n}$. Note that this is a non-linear combination of the linear predictors η_i and that, in addition, the total effects are non-linear on $\overline{f(\eta)}$ and β_r. Hence, the total effects cannot be computed using INLA for a Probit model. The same restriction also applies to the direct effects, which depend on $D(f(\eta))$, a non-linear function of η.

An approximation may be computed by replacing each η_i by $\hat{\eta}_i = E[\eta_i|y]$, so that $D(f(\eta))$ is replaced by $D(f(\hat{\eta}))$. Note that this will require fitting the model twice. Firstly, to obtain $\hat{\eta}_i$ and, secondly, to include $D(f(\hat{\eta}))$ in the computation of the effects and impacts. Looking at equations (22), (23) and (24), we can see that taking $D(f(\hat{\eta}))$ will make the impacts a linear combination on β_r and, possibly, γ_k but will ignore the uncertainty about η.

Table 1: Different types of impacts for the different spatial econometrics models assuming a Gaussian response.

Model	$S_r(W)$	$\overline{M}(r)_{direct}$	$\overline{M}(r)_{total}$
SEM	$I_n \beta_r$	β_r	β_r
SLM	$(I_n - \rho_{\text{Lag}} W)^{-1} \beta_r$	$n^{-1} tr\Big((I_n - \rho_{\text{Lag}} W)^{-1}\Big)\beta_r$	$\frac{1}{1-\rho_{\text{Lag}}}\beta_r$
SDM	$(I_n - \rho_{\text{Lag}} W)^{-1}(\beta_r + W\gamma_r)$	$n^{-1} tr\Big((I_n - \rho_{\text{Lag}} W)^{-1}\Big)\beta_r +$ $+ n^{-1} tr\Big((I_n - \rho_{\text{Lag}} W)^{-1} W\Big)\gamma_r$	$\frac{1}{1-\rho_{\text{Lag}}}(\beta_r + \gamma_r)$

Table 2: Different types of impacts for the different Spatial Econometrics models assuming a Binary response.

Model	$S_r(W)$	$\overline{M}(r)_{direct}$	$\overline{M}(r)_{total}$
SEM	$D(f(\eta))\beta_r$	$\beta_r \sum_{i=1}^{n} \frac{f(\eta_i)}{n}$	$\beta_r \sum_{i=1}^{n} \frac{f(\eta_i)}{n}$
SLM	$D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1}\beta_r$	$n^{-1} tr\Big(D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1}\Big)\beta_r$	$\frac{\beta_r}{(1-\rho_{\text{Lag}})} \sum_{i=1}^{n} \frac{f(\eta_i)}{n}$
SDM	$D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1}(\beta_r + W\gamma_r)$	$n^{-1} tr\Big(D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1}\Big)\beta_r +$ $n^{-1} tr\Big(D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1} W\Big)\gamma_r$	$\frac{\beta_r + \gamma_r}{(1-\rho_{\text{Lag}})} \sum_{i=1}^{n} \frac{f(\eta_i)}{n}$

3. Approximate Inference using the Integrated Laplace Approximation

Rue et al. [23] and Lindgren et al. [15] have developed an approximate method for Bayesian inference based on focusing on the marginals of the parameters of the models. They consider the class of Latent Gaussian Markov Random Fields, which are flexible enough to be used in many different types of applications.

Briefly, given a vector of observed variables $\mathbf{y} = (y_1, \ldots, y_n)$, the distribution of y_i is assumed to belong to the exponential family with mean μ_i. The relation between μ_i and a linear predictor of some latent effects is expressed through a link function. The linear predictor η_i can include fixed and random effects as well as other non-linear terms on some covariates. The distribution of $\mathbf{y}$ may also depend on a vector of hyperparameters θ_1.

The vector of latent effects $\mathbf{x}$ (which will include the ensemble linear predictor for each observation, coefficients of the covariates, etc.) is assumed to be a Gaussian Markov Random Field (GMRF) with precision matrix $Q(\theta_2)$, where θ_2 is a vector of hyper-parameters.

Hence, the observations are independent given $\mathbf{x}$ and $\theta = (\theta_1, \theta_2)$ and the model likelihood can be written down as

$$\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i \in \mathcal{I}} \pi(y_i | x_i, \theta), \tag{31}$$

where x_i is the latent linear predictor η_i and $\mathcal{I}$ represents the indices of the observations. Note that some of the values in $\mathbf{y}$ may be missing and this is why the product is over a set of indices $\mathcal{I}$ and not from 1 to n.

INLA will provide accurate approximations to the posterior marginals of the model parameters and hyper-parameters. These approximations are based on providing a multidimensional integration of all the other latent effects and hyperparameters. For example, the joint distribution of the model parameters and hyperparameters is:

$$\pi(\mathbf{x}, \theta|\mathbf{y}) \propto \pi(\theta)\pi(\mathbf{x}|\theta) \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \theta) \propto \tag{32}$$

$$\propto \pi(\theta)|\mathbf{Q}(\theta)|^{n/2} \exp\{-\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\theta)\mathbf{x} + \sum_{i \in \mathcal{I}} \log(\pi(y_i|x_i, \theta)\}.$$

Note that the expression $|\mathbf{Q}(\theta)|$ will include what is referred to as the 'Jacobian' (i.e., $I_n - \rho W$) in the spatial econometrics literature [see, for example, 1].

The marginal distributions for the latent effects and hyper-parameters can be written as

$$\pi(x_i|\mathbf{y}) \propto \int \pi(x_i|\theta, \mathbf{y})\pi(\theta|\mathbf{y})d\theta, \tag{33}$$

and

$$\pi(\theta_j|\mathbf{y}) \propto \int \pi(\theta|\mathbf{y}) d\theta_{-j}. \quad (34)$$

Rue et al. [23] provide a simple approximation to $\pi(\theta|\mathbf{y})$, denoted by $\tilde{\pi}(\theta|\mathbf{y})$, which is then used to compute the approximate marginal distribution of a latent parameter x_i:

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\theta_k, \mathbf{y}) \times \tilde{\pi}(\theta_k|\mathbf{y}) \times \Delta_k, \quad (35)$$

where Δ_k are the weights of a particular vector of values θ_k in a grid for the ensemble of hyperparameters.

Rue et al. [23] also discuss how the approximation $\tilde{\pi}(x_i|\theta_k, \mathbf{y})$ should be in order to reduce numerical error and they provide different alternatives. Finally, an R [19] package called **R-INLA** is available to fit a large range of models using the Integrated Nested Laplace Approximation. For a general discussion on the applications of INLA in spatial statistics see Gómez-Rubio et al. [6].

4. Model fitting with INLA

At the moment, the **R-INLA** software cannot fit the class of models described in Sections 2.1 and 2.2. First of all, the SAR specification for the variance-covariance matrix is not available yet. Secondly, the linear term on the covariates in the Spatial Lag and Spatial Durbin models is multiplied by $(I - \rho W)^{-1}$, which is clearly not a standard linear term. Note that in this section we will use ρ for the spatial autocorrelation parameter regardless of the model used.

After conditioning on a value of ρ, these models belong to the class of models that **R-INLA** can fit and the posterior marginals are also conditioned on the value of ρ. Hence, for a given $\rho = \rho_0$ the likelihood will become $\pi(y|\theta, \rho_0)$ and hence INLA will provide approximations to the (conditioned) marginals

$$\pi(x_i|y, \rho_0), \quad (36)$$

and the marginal likelihoods reported by INLA are also conditioned on the value of ρ, i.e., we obtain $\pi(y|\rho_0)$ instead of $\pi(y)$. The marginal distribution of ρ is

$$\pi(\rho|y) = \frac{\pi(y|\rho)\pi(\rho)}{\pi(y)} \propto \pi(y|\rho)\pi(\rho), \quad (37)$$

where $\pi(\rho)$ represents a prior distribution for ρ.

As discussed in Section 2.1, ρ is restricted to a bounded interval. We can take a fine one-dimensional grid in this interval $\{\rho_j\}_{j=1}^r$, so that for each value we can fit a different model and use all the information reported to approximate the marginal distribution of ρ. In practice, this interval will depend on the domain of the prior distribution on ρ. For example, when a uniform prior on (0,1) is used, the grid will be defined in this interval.

For each value of ρ_j we can compute $\pi(y|\rho_j)\pi(\rho_j)$, a value that is proportional to the actual values of $\pi(\rho_j|y)$. $\pi(\rho|y)$ can be obtained by fitting a curve (for example, using splines) to points $\{(\rho_j, \pi(y|\rho_j)\pi(\rho_j))\}_{j=1}^{r}$ and re-scaling it to integrate one.

Note that in our case, the values of the parameter of interest are bounded and this makes the approach easier. We may be interested in applying this approach to other models where we would like to compute the marginal distribution of a parameter which is not bounded. In this case, we may proceed in a different way to make computations more efficiently.

First of all, if our parameter is, say, λ, we could use an optimisation algorithm to find the maximum (mode) of $\pi(y|\lambda)$. This requires the evaluation of several models for different values of λ. Once the mode has been obtained, an interval around the mode can be set so that the difference in the marginal log-likelihoods between the mode and the interval limits is large (for example, 10). This can be easily implemented as well. At this stage, the interval can be divided using a one-dimensional grid $\{\lambda_j\}_{j=1}^{l}$, computing $\pi(y|\lambda_j)\pi(\lambda_j)$, fitting a curve to these values and then obtaining $\pi(\lambda|y)$ by re-scaling the curve to integrate one.

Although this approach may seem computationally intensive and not providing fast results, most computations can be done in parallel. In particular, the models that arise from different values of λ can be fitted on separate nodes of a cluster. Although MCMC methods can run several chains in parallel, computation for single chains is often impossible to run on several nodes as future samples depend on current values.

Furthermore, this approach could be used to compute the posterior distribution of pairs of parameters. For example, in the Spatial Durbin Model we may be interested in any possible interaction between the spatial autocorrelation parameter ρ and the coefficient of one of the lagged covariates γ_l. For a fixed values of ρ and γ_l $(\rho_0, \gamma_{l,0})$ the model can be fitted to obtain $\pi(y|\rho_0, \gamma_{l,0})$.

Again, we can create a grid of values $\{\{(\rho_j, \gamma_{l,k})\}_{j=1}^{r}\}_{k=1}^{g}$ (in two dimensions now) to compute the marginal likelihood given a pair of values for ρ and γ_l. Now the (bivariate) posterior distribution of the pair (ρ, γ_l) is given by:

$$\pi(\rho, \gamma_l|y) \propto \pi(y|\rho, \gamma_l)\pi(\rho, \gamma_l). \tag{38}$$

A convenient way of taking a prior for ρ and γ_l is

$$\pi(\rho, \gamma_l) = \pi(\rho)\pi(\gamma_l). \tag{39}$$

Hence, the posterior distribution can be computed by fitting a surface to points

$$\Big(\rho_j,\ \gamma_{l,k},\ \pi(y|\rho_j, \gamma_{l,k})\pi(\rho_j, \gamma_{l,k})\Big);\ j = 1, \ldots, r;\ k = 1, \ldots, g\ , \tag{40}$$

and re-scaling it to integrate one.

From this bivariate distribution credible regions can be computed and correlation between ρ and γ_l can be assessed. Again, this procedure can be easily parallelized to reduce computational time. Note also that accuracy can be improved by defining a less coarse grid.

Extending this idea to more than two variables is easy, but the computational burden increases as well and at some point it may be preferable to fit the models using MCMC methods. Finally, we would like to mention that INLA can be integrated into more general MCMC algorithms to integrate parameters out at some stages. For example, when fitting a model using Reversible-jump, INLA could be used to fit the resulting model after the dimension of the model has been sampled. How all the resulting models can be combined is explained in Section 4.2. Using INLA within MCMC might need further assumptions, as we compute marginals but the joint is required. So unless a Gaussian (or a mixture thereof) is used, errors will be introduced.

4.1. Bayesian Model Selection

When making inference on ρ, it can be regarded as a truly discrete variable and we may be interested in estimating the posterior probabilities for each value. Hence, we will try to estimate a probability function. Each value of ρ will produce a slightly different model, so in the end we are performing a model selection.

Suppose that we have a set of models $\{\mathcal{M}_j\}_{j=1}^r$, each one associated to a value of $\rho = \rho_j$. First, we need to assign a prior distribution to each model. A uniform distribution can be used as a vague prior to give the same prior probability to each model, i.e., $\pi(\mathcal{M}_i) = 1/r,\ \forall j = 1, \ldots, r$. Hence, $\pi(y|\mathcal{M}_i) \equiv \pi(y|\rho = \rho_j)$.

The posterior probability for each model can be computed as

$$\pi(\mathcal{M}_i|y) \propto \pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i). \tag{41}$$

Given that now the number of models proposed is r, the posterior probabilities can be re-scaled to integrate one dividing by

$$\sum_{j=1}^{r} \pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i), \tag{42}$$

so that

$$\pi(\mathcal{M}_i|y) = \frac{\pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i)}{\sum_{j=1}^{r} \pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i)}. \tag{43}$$

When a uniform prior is assigned to the models the previous expression simplifies to

$$\pi(\mathcal{M}_i|y) = \frac{\pi(y|\mathcal{M}_i)}{\sum_{j=1}^{r} \pi(y|\mathcal{M}_i)}. \tag{44}$$

Then, inference on the other parameters can be based upon the model with the highest posterior probability with no need of model averaging. This is an alternative approach when the computational burden of estimating the posterior distribution of ρ is too high.

LeSage and Pace [13] use MCMC to estimate the posterior distributions of the model parameters and, in particular, a griddy Gibbs sampler [21] for ρ. In the next section we propose the use of Bayesian Model Averaging to estimate the posterior marginal of ρ and all the other model parameters.

4.2. Bayesian Model Averaging with INLA

So far, we have explained how **R-INLA** can be used to obtain an approximation to $\pi(\rho|y)$ and $\pi(\rho, \gamma_l|y)$ even if our model is not implemented in the software. In order to obtain the marginal distributions of the other parameters in the model, the parameters we are conditioning on need to be integrated out as follows. For the unidimensional case with ρ, this proceeds as follows:

$$\pi(x_i|y) = \int \pi(x_i|\rho, y)\pi(\rho|y)d\rho = \int \pi(x_i|\rho, y)\frac{\pi(y|\rho)\pi(\rho)}{\pi(y)}d\rho. \tag{45}$$

This integral can be approximated using a grid on the values of ρ:

$$\pi(x_i|y) \simeq \sum_{j=1}^{r} \pi(x_i|\rho_j, y)\pi(\rho_j|y)\Delta_j, \tag{46}$$

where Δ_j are weights which, in the simplest case, are equal to the size of the intervals in the grid. If our grid is equally spaced, then $\Delta_j = \Delta \; \forall j$.

Noting that

$$\pi(y) = \int \pi(y|\rho)\pi(\rho)d\rho \simeq \sum_{j=1}^{r} \pi(y|\rho_j)\pi(\rho_j)\Delta_j, \tag{47}$$

the approximation can be written as

$$\pi(x_i|y) \simeq \sum_{j=1}^{r} \pi(x_i|\rho_j, y)\frac{\pi(y|\rho_j)\pi(\rho_j)}{\sum_{j'=1}^{r} \pi(y|\rho_{j'})\pi(\rho_{j'})\Delta_{j'}}\Delta_j. \tag{48}$$

For equally spaced grids, where $\Delta_j = \Delta_{j'} = \Delta$ the previous expression simplifies, so that:

$$\pi(x_i|y) \simeq \sum_{j=1}^{r} \pi(x_i|\rho_j, y)\frac{\pi(y|\rho_j)\pi(\rho_j)}{\sum_{j'=1}^{r} \pi(y|\rho_{j'})\pi(\rho_{j'})}. \tag{49}$$

Alternatively, this can be expressed as a weighted sum of the marginal distributions provided by each of the fitted models (for a given value of ρ):

$$\pi(x_i|y) \simeq \sum_{j=1}^{r} \pi(x_i|\rho_j, y)\lambda_j; \;\; \lambda_j = \frac{\pi(y|\rho_j)\pi(\rho_j)}{\sum_{j'=1}^{r} \pi(y|\rho_{j'})\pi(\rho_{j'})}. \tag{50}$$

Hence, marginal inference is still possible with this Bayesian Model Averaging approach [12]. Note that weights only depend on the marginal likelihood and the prior of ρ and not on the parameter of interest x_i. Furthermore, this procedure can be easily parallelised to reduce computational time.

Summary statistics for the distribution of x_i can be easily derived as well using this approach. For example:

$$E[x_i|y] \simeq \sum_{j=1}^{r} E[x_i|\rho_j, y]\lambda_j. \quad (51)$$

That is, the posterior mean of x_i is a weighted sum of the different posterior means computed by conditioning on different values of ρ. If vague priors for the other parameters in the model are used, the posterior modes should be close to maximum likelihood estimates.

4.3. Computing the impacts with INLA

In order to compute the distribution of the impacts with INLA it si clear that that for the Gaussian case, and conditioning on ρ, the impacts are a linear combination of β_k and, possibly, γ_k. These linear combinations can easily be computed with INLA and then the final distribution of the impacts can be computed by Bayesian Model Averaging.

As already discussed in Section 2.4, for the Probit models the impacts also depend on $D(f(\eta))$ which has its own posterior distribution. An approximation to the posterior distribution of the impacts can be provided by using $D(f(\hat{\eta}))$, where $f(\hat{\eta})$ is the vector of the posterior mean of $f(\eta)$, so that the impacts become a linear combination of β_k and, possibly, γ_k.

4.4. INLA versus MCMC

LeSage and Pace [13] describe a number of MCMC samplers to fit the different models described therein. In general, they implement Gibbs and Metropolis-Hastings samplers for the spatial autocorrelation parameter. In both cases, the value of the determinant of the Jacobian (denoted by $|I - \rho W|$) is required for many different values of ρ which are not known beforehand. As this is computationally expensive, LeSage and Pace [13] use the following procedure.

A grid of values in its domain is defined so that $|I - \rho W|$ is precomputed for each value of ρ in the grid. For values of ρ not in the grid, the value of $|I - \rho W|$ is interpolated (using a spline, for example). This computational approach is important as computing the Jacobian can be computationally expensive when large matrices are involved. Furthermore, the grid of values of ρ needs to be thin, with a step size of 0.01.

The (approximated) value of $|I - \rho W|$ is later used to compute the different conditional distributions that appear in the griddy Gibbs sampling or the ratio in the Metropolis-Hastings algorithm. This means that a new interpolated value needs to be obtained at each step of the MCMC algorithm.

Note that the approach described in this paper also relies on fitting models given different values of ρ in a grid. This reduces the computational burden as well. Griffith [8] describes a different approach to estimate the Jacobian which is also computationally efficient.

Our approach also requires a good approximation to the marginal likelihood of the different values fitted given ρ. R-INLA provides a good approximation to the marginal likelihood. The marginal likelihood should vary smoothly for slighty different values of ρ, which in turn will produce a smooth posterior distribution for ρ. We have noticed that in some cases the values of the marginal likelihood may wiggle (for models associated to similar values of ρ).

This is not due to errors in the computation of the marginal likelihood but to differences in the estimation of the hyperparameters. In general, for two different but similar values of ρ we have obtained similar posterior marginals of the fixed effects but slightly different posterior marginals of the hyperparameters (namely, the precission of the spatial effects). In any case, we believe that this is not a problem as the marginal likelihood is estimated with accuracy and poor estimates of the distribution of the hyperparameters will lead to lower values of the marginal likelihood.

Finally, we have not explored the possibility that the marginal likelihood is multimodal, as it is known to happen to the likelihood for some spatial models [see 20, 16]. This is not a problem per se in our examples but it may require a finer grid on the values of ρ to capture this variability.

5. Simulation study

In order to assess the differences between the models we have conducted a simulation study. We are interested in how data generated under one model are fitted by other models. This will provide some insight into the adequacy of the different specifications for the spatial structure in the data.

To simulate the datasets, we have considered the Gaussian and Binomial cases. In the Gaussian case, we have simulated 500 datasets considering a model with a single covariate and proper CAR random effect with variance-covariance matrix $\Sigma = \sigma^2(I_n - \frac{\rho}{\lambda_{max}}W)^{-1}$. This model can be written down as

$$y_i = \beta x_i + u_i;\ u_i \sim N(0, \sigma^2(I_n - \frac{\rho}{\lambda_{max}}W)^{-1}); i = 1, \dots, 100. \qquad (52)$$

β has been set to 1, ρ to 0.5 and σ to 3. In this case, W is a binary and symmetric adjacency matrix and λ_{max} is the maximum eigenvalue of W. In order to develop the spatial structure, we consider observations to be on a straight line, so that neighbours are the observations to the left and right. The edges are considered as neighbours too. In practice, the spatial structure is as if observations are located on a circle and all locations have 2 neighbors. Hence, the adjacency matrices for the CAR and SAR specifications are the same except for a scaling factor of exactly 1/2.

Furthermore, we have simulated another 100 datasets according to each of the 3 spatial econometric models described in Section 2.1. The parameters are the same as in the previous model. For the Spatial Durbin Model the coefficient of the lagged covariates has been set to 1. For these models, the weight matrix W is row-standardised.

Table 3: Relative Root Mean Square Error of different Gaussian models fitted to 4 × 500 datasets simulated under those models.

	Model			
	CAR	SEM	SLM	SDM
data CAR	3.330101e-11	3.319643e-11	3.531584e-11	3.539592e-11
data SEM	4.654049e-11	5.103125e-11	4.775117e-11	5.120926e-11
data SLM	4.746212e-11	4.759239e-11	4.703661e-11	4.686431e-11
data SDM	6.077251e-11	6.677201e-11	7.368624e-11	7.650566e-11

Hence, for the Gaussian case we have 4 × 500 different datasets generated under four different models. Our aim is to assess how good are predictions from one model on the different datasets.

For the Binomial case we have simulated 400 datasets in a similar way. In this case, a latent variable y_i^* has been simulated in the same way as in the Gaussian case. Then it has been transformed according to a probit link and the resulting probability rounded to 0 or 1.

For each dataset and model fit we will consider the Relative Root Mean Square Error (RRMSE) as a measure of goodness-of-fit:

$$RRMSE = \sqrt{\frac{1}{100}\sum_{i=1}^{100}\frac{(y_i - \hat{y}_i)^2}{y_i^2}}. \tag{53}$$

In the case of the Binomial data, we have computed the average of the rightly classified observations.

5.1. Gaussian models

Table 3 shows the averages of the RRMSE for each type of dataset and the models fitted. In general, all of them provide similar fitting, and being good in all cases.

5.2. Non-Gaussian models

Table 4 shows the averages of the proportion of correctly classified observations. We have also included results from a naïve model with no spatial structure at all for reference. All models predict with an accuracy higher than an 58%. The SEM model seems to be the one with the worst performance in all cases, being close to the naïve model. The CAR model shows the highest prediction rate, except for data generated under the SDM model. The SLM and SDM models show a good performance in all cases.

5.3. Implementation issues

For the implementation of the methods described in this paper we have made extensive use of the R programming language and the **R-INLA** package. The methods described in this paper have been implemented in package **INLABMA**,

Table 4: Percentage of right classified using different Probit models to different datasets simulated under those models.

	Model				
	CAR	SEM	SLM	SDM	Naïve
data CAR	62.44	58.29	58.45	59.47	58.28
data SEM	60.53	58.41	58.77	60.31	58.41
data SLM	64.97	60.96	61.31	61.97	60.96
data SDM	66.77	63.29	64.58	68.02	63.29

which is available from CRAN, and the full code used in this paper is available from the authors on request. We have also tested our methods on a small example with the proper CAR model that **R-INLA** can fit to make sure that we obtained the same results. This has been done for the Gaussian and non-Gaussian cases and we found that our code fitted the results provided by INLA quite well. Finally, we have double-checked with an R implementation of some of the functions in the Spatial Econometrics toolbox.

We have also observed that thinning was often required to obtain the marginal distributions with MCMC, and that it was not clear how many iterations where required beforehand. This is not a problem using our approach. Another advantage that we have found in using **R-INLA** is that it is easier to use different priors on the model parameters. In particular, the number of available priors in **R-INLA** is larger. This allows us to assess the impact of priors on the results, or choose informative priors when there is not enough information in the data.

Furthermore, **R-INLA** provides other latent effects not discussed in this paper that may be used with our current implementation. These effects may be used, for example, to include temporal or longitudinal effects in the data. Our code and approach will still be valid when other latent effects are included in the model using a standard call to the **R-INLA** function for model fitting `inla()`.

Finally, INLA and **R-INLA** can handle missing data to provide predictive distributions. This means that the uncertainty about the missing data is taken into account when computing the marginals of the model parameters and that we can compute summary statistics on the missing values. Needless to say, when using spatial models, spatial dependence will help to obtain better predictions of the missing values.

6. Examples

6.1. Boston housing data

In this example we re-analyse the Boston housing data described by Harrison and Rubinfeld [11]. Here the interest is in estimating the median value of owner-occupied houses (which has been censored at $50k) using 13 relevant covariates and the spatial structure of the data Pace and Gilley [18]. The data is available in the R package **spdep** (`boston` dataset), that also includes methods for maximum likelihood estimation of the models discussed in this paper. We

Table 5: Point estimates of the fixed effects for the Boston housing data using different models.

	MLCAR	INLACAR	MLSEM	INLASEM	MLSLM	INLASLM
(Intercept)	4.11e+00	4.116596	3.84e+00	3.923540	2.279626	2.308686
CRIM	-8.14e-03	-0.008197	-5.29e-03	-0.005886	-0.007105	-0.007267
ZN	1.56e-04	0.000150	4.73e-04	0.000372	0.000380	0.000365
INDUS	-4.87e-05	-0.000019	-2.52e-05	0.000158	0.001257	0.001376
CHAS1	3.18e-02	0.033327	-3.88e-02	-0.008519	0.007368	0.012609
I(NOX^2)	-4.61e-01	-0.467009	-2.23e-01	-0.339600	-0.268916	-0.286864
I(RM^2)	8.08e-03	0.008042	7.96e-03	0.008052	0.006724	0.006763
AGE	-4.88e-04	-0.000478	-1.05e-03	-0.000787	-0.000277	-0.000226
log(DIS)	-1.64e-01	-0.164862	-1.18e-01	-0.136482	-0.158301	-0.157636
log(RAD)	7.83e-02	0.078420	6.55e-02	0.074014	0.070689	0.073167
TAX	-3.97e-04	-0.000399	-5.00e-04	-0.000475	-0.000366	-0.000364
PTRATIO	-2.10e-02	-0.021258	-1.77e-02	-0.020105	-0.012011	-0.012743
B	5.07e-04	0.000507	5.94e-04	0.000585	0.000284	0.000289
log(LSTAT)	-3.21e-01	-0.321995	-2.66e-01	-0.276658	-0.232161	-0.232201

have fitted the 3 spatial econometrics models described in this paper plus a spatial model with a CAR error term. In addition, we have fitted the spatial econometrics models using maximum likelihood to compare the estimates of the model parameters. Adjacency is defined in the **R** object `boston.soi` and a binary weight matrix has been used to fit the CAR model whilst a row standardised matrix has been used for all the other models.

The results are summarised in Tables 5, 6 and 7. In general, Bayesian and maximum likelihood estimates are close. For the Bayesian models we also have the posterior marginals that allow us to compute exact credible intervals for the covariate coefficients and the spatial autocorrelation parameter ρ.

Furthermore, we have plotted the posterior marginal distribution of ρ under different models in Figure 1. The different posterior distributions should not be surprising as the models are in fact different and they have different spatial correlation structures. Table 7 summarises the estimates of the spatial autocorrelation parameter ρ. Again, the Bayesian estimates are close to those obtained by maximum likelihood. The actual estimates for the CAR models with INLA are at a different scale than the maximum likelihood ones, because INLA re-scales the adjacency matrix by its maximum eigenvalue (same as the proper CAR model described in Section 5), which is 5.306. Hence, the values reported in Table 7 are the original values divided by 5.306 so that they can be compared to the maximum likelihood estimates. Taking this into account, we notice that both estimates are close to each other. Also, estimates for the maximum likelihood CAR and SEM models have been computed by resampling from the fitted models.

Impacts are displayed in Tables 8, 9 and 10. Although only mean and standard deviation are displayed it is possible to obtain credible intervals and other statistics from the posterior marginal distributions of the impacts. These

Table 6: Point estimates of the fixed effects for the Boston housing data using a Spatial Durbin Model.

	MLSDM	INLASDM	lag-MLSDM	lag-INLASDM
(Intercept)	1.898178	1.998026	—	—
CRIM	-0.005710	-0.005833	-0.004642	-0.004727
ZN	0.000691	0.000630	-0.000379	-0.000256
INDUS	-0.001113	-0.001495	0.000251	0.001274
CHAS1	-0.041632	-0.028905	0.125183	0.100878
I(NOXˆ2)	-0.010349	-0.006873	-0.386407	-0.418873
I(RMˆ2)	0.007950	0.007825	-0.004513	-0.004206
AGE	-0.001288	-0.001214	0.001497	0.001523
log(DIS)	-0.124041	-0.122184	-0.004539	-0.011101
log(RAD)	0.058635	0.055938	-0.009407	0.003441
TAX	-0.000491	-0.000474	0.000411	0.000350
PTRATIO	-0.013199	-0.013531	0.000603	-0.000595
B	0.000564	0.000540	-0.000508	-0.000429
log(LSTAT)	-0.247245	-0.249890	0.098467	0.091380

Table 7: Estimates of the spatial autocorrelation parameter for the different models.

	rho	rhosd	LLrho	ULrho
MLCAR	0.186	0.0039	0.178	0.194
INLACAR	0.183	0.0024	0.177	0.186
MLSEM	0.708	0.0335	0.642	0.774
INLASEM	0.687	0.0318	0.623	0.749
MLSLM	0.485	0.0294	0.428	0.543
INLASLM	0.477	0.0274	0.423	0.529
MLSDM	0.596	0.0384	0.520	0.671
INLASDM	0.571	0.0378	0.496	0.643

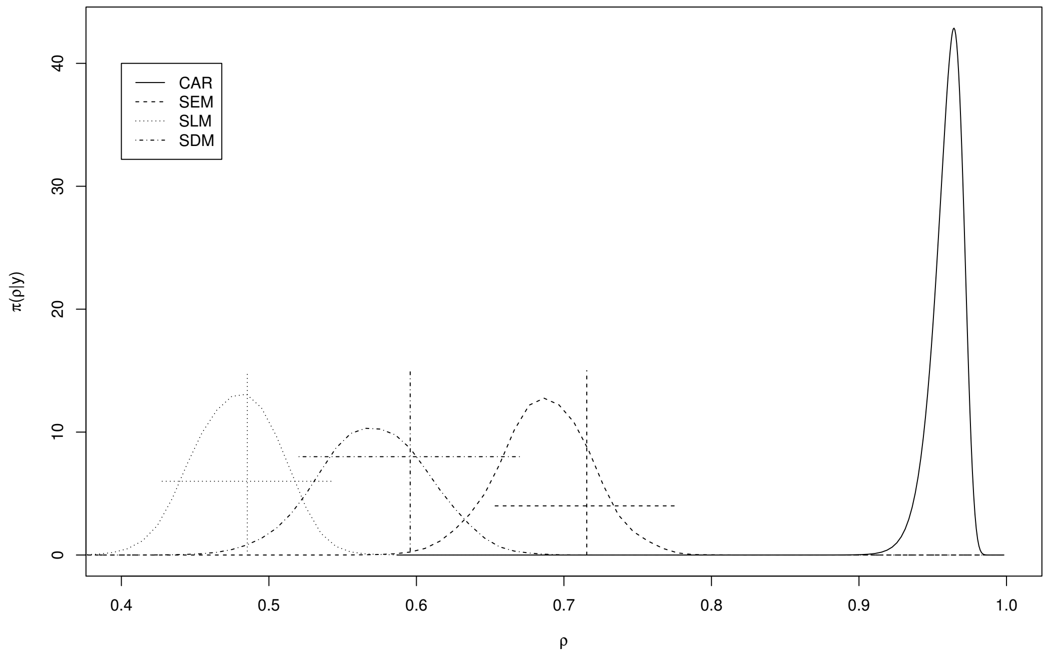

Figure 1: Marginal distributions for the spatial autocorrelation parameter ρ under different models (Boston housing data). Lines show ML point estimates with approximate 95% confidence intervals.

can be useful to assess significance of the impacts.

6.2. After-Katrina business data

In this example we consider the data analysed in LeSage et al. [14] regarding the probability of re-opening a business in the aftermath of hurricane Katrina. In this case we have a non-Gaussian model because we are modelling a probability and the response variable can take either 1 (the business re-opened) or 0 (the business didn't re-open). As in the previous example, we have fitted four models, but we have now used a GLM with a Binomial family and a probit link. Adjacencies have been obtained using k-nearest neighbours (with $k = 11$) on the coordinates of the businesses. A binary weight matrix has been used to fit the CAR model, whilst a row standardised matrix have been used to fit the remaining models.

LeSage et al. [14] split the data into four periods according to different time frames. In our analysis we focus on the first period, i.e., the business re-opened during the first 3 months (90 days). The model used there is the one that we

Table 8: Direct impacts for the Boston housing data. Impacts for the SEM model are not displayed as they are equal to the point estimes of the fixed effects.

	mean (SLM)	sd (SLM)	mean (SDM)	sd (SDM)
CRIM	-0.007791	1.04e-03	-0.007454	0.001008
ZN	0.000393	4.12e-04	0.000651	0.000508
INDUS	0.001481	1.91e-03	-0.001407	0.002915
CHAS1	0.013343	2.72e-02	-0.012034	0.028185
I(NOXˆ2)	-0.307155	9.26e-02	-0.091868	0.183783
I(RMˆ2)	0.007262	1.08e-03	0.007883	0.001086
AGE	-0.000243	4.29e-04	-0.001048	0.000485
log(DIS)	-0.169146	2.69e-02	-0.138429	0.089886
log(RAD)	0.078528	1.53e-02	0.063073	0.021372
TAX	-0.000391	9.94e-05	-0.000458	0.000119
PTRATIO	-0.013642	4.12e-03	-0.015206	0.005691
B	0.000310	8.40e-05	0.000516	0.000108
log(LSTAT)	-0.248953	2.12e-02	-0.260125	0.022665

Table 9: Total impacts for the Boston housing data. Impacts for the SEM model are not displayed as they are equal to the point estimes of the fixed effects.

	mean (SLM)	sd (SLM)	mean (SDM)	sd (SDM)
CRIM	-0.013923	0.001837	-0.024623	0.003721
ZN	0.000703	0.000739	0.000872	0.001154
INDUS	0.002656	0.003428	-0.000485	0.005003
CHAS1	0.023598	0.048572	0.167016	0.084658
I(NOXˆ2)	-0.548308	0.163024	-0.992875	0.217646
I(RMˆ2)	0.012993	0.002043	0.008482	0.003240
AGE	-0.000436	0.000768	0.000721	0.001216
log(DIS)	-0.302466	0.049318	-0.310654	0.069526
log(RAD)	0.140451	0.027923	0.138643	0.044279
TAX	-0.000700	0.000180	-0.000289	0.000305
PTRATIO	-0.024347	0.007237	-0.032946	0.011023
B	0.000554	0.000152	0.000261	0.000215
log(LSTAT)	-0.444868	0.037011	-0.368753	0.057901

Table 10: Indirect impacts for the Boston housing data. Impacts for the SEM model are not displayed as they are all equal to zero.

	mean (SLM)	sd (SLM)	mean (SDM)	sd (SDM)
CRIM	-0.006119	9.05e-04	-0.017169	0.003303
ZN	0.000310	3.28e-04	0.000221	0.001140
INDUS	0.001173	1.52e-03	0.000922	0.005570
CHAS1	0.010220	2.14e-02	0.179050	0.077119
I(NOXˆ2)	-0.240802	7.25e-02	-0.901008	0.276120
I(RMˆ2)	0.005711	1.02e-03	0.000599	0.002825
AGE	-0.000193	3.39e-04	0.001769	0.001164
log(DIS)	-0.133018	2.41e-02	-0.172225	0.110585
log(RAD)	0.061787	1.33e-02	0.075570	0.047068
TAX	-0.000308	8.37e-05	0.000169	0.000301
PTRATIO	-0.010690	3.21e-03	-0.017740	0.011819
B	0.000244	6.97e-05	-0.000255	0.000216
log(LSTAT)	-0.195394	2.03e-02	-0.108629	0.054616

Table 11: Point estimates of the fixed effects for the Katrina dataset using different models.

	INLACAR	INLASEM	INLASLM	INLASDM
(Intercept)	-16.926	-18.599	-8.067	-10.326
flood	-0.446	-0.459	-0.204	-0.476
log(medinc)	1.646	1.809	0.767	1.948
sizeempsmall	-0.376	-0.364	-0.423	-0.420
sizeemplarge	-0.402	-0.389	-0.441	-0.340
sesstatuslow	-0.482	-0.348	-0.502	-0.025
sesstatushigh	0.110	0.111	0.085	0.035
owntypesole	0.837	0.820	0.854	0.903
owntypenational	0.089	0.102	0.166	0.130

have termed the Spatial Lag Model in this paper. For comparison purposes, the model that they have termed SAR (spatial auto-regressive) is our SLM model. Following LeSage and Pace [13, page 283] we have also set $\sigma_e^2 = 1$ because of the identifiability problem between β and σ_e^2.

Table 11 shows the point estimates of the coefficient covariates used in the model. In this case we do not report maximum likelihood estimates because there is no similar code available in **spdep**. The `spprobitml` in the **McSpatial** package re-casts the spatial weights matrix as block-diagonal, and thus does not estimate the same model because a different weight matrix is used; the `spprobit` function in the same package fits a linearized GMM model. The coefficient estimates are close to those obtained by fitting a standard probit model with the `glm()` function in R. In addition, our estimates for the SLM model are similar to those reported in LeSage et al. [14].

Regarding spatial autocorrelation, Table 12 shows summary statistics of ρ

Table 12: Estimates of the spatial autocorrelation parameter for the different models on the Katrina dataset.

	rho	rhosd	LLrho	ULrho
INLACAR	0.836	0.135	0.474	0.983
INLASEM	0.597	0.105	0.345	0.756
INLASLM	0.495	0.107	0.251	0.669
INLASDM	0.214	0.144	0.018	0.521

Table 13: Direct impacts for the SLM model (after-Katrina business data).

	Mean	Lower 0.05	Upper 0.95
flood	-0.044	-0.065	-0.027
log(medinc)	0.167	0.063	0.281
sizeempsmall	-0.093	-0.162	-0.026
sizeemplarge	-0.097	-0.258	0.060
sesstatuslow	-0.111	-0.187	-0.037
sesstatushigh	0.019	-0.044	0.081
owntypesole	0.189	0.095	0.284
owntypenational	0.037	-0.145	0.215

under different models. We have found a small difference between our estimate using the SLM model and the one reported by LeSage et al. [14]. The posterior marginals have been plotted in Figure 2. Note the SDM model seems to favour a null value of ρ meaning that once we have included the lagged covariates there is no positive spatial autocorrelation left.

As discussed in Section 2.4, impacts cannot be computed in an exact way but we believe that a reasonable approximation can be obtained. We have included direct, total and indirect and total impacts for the SLM model in Tables 13, 14 and 15, respectively. Point estimates and credible intervals are often similar to those reported for the SAR model. Not having a perfect match should not be surprising considering how we have approximated the posterior marginals of the impacts. If credible intervals are used to assess significance of the impacts, however then we obtained the same conclusions as in LeSage et al. [14].

7. Other models and applications in Spatial Econometrics

In this paper we have considered a reduced family of the Spatial Econometrics models described in LeSage and Pace [13]. They describe a good number of other spatial models that could benefit from INLA, such as models with different spatial scales and techniques for model selection.

The SLM model includes a single weight matrix but LeSage and Pace [13, page 150] discuss a model in which there are two different weight matrices to model autocorrelation on the response and autocorrelation on the error term. They have termed this model SAC and it is defined as follows:

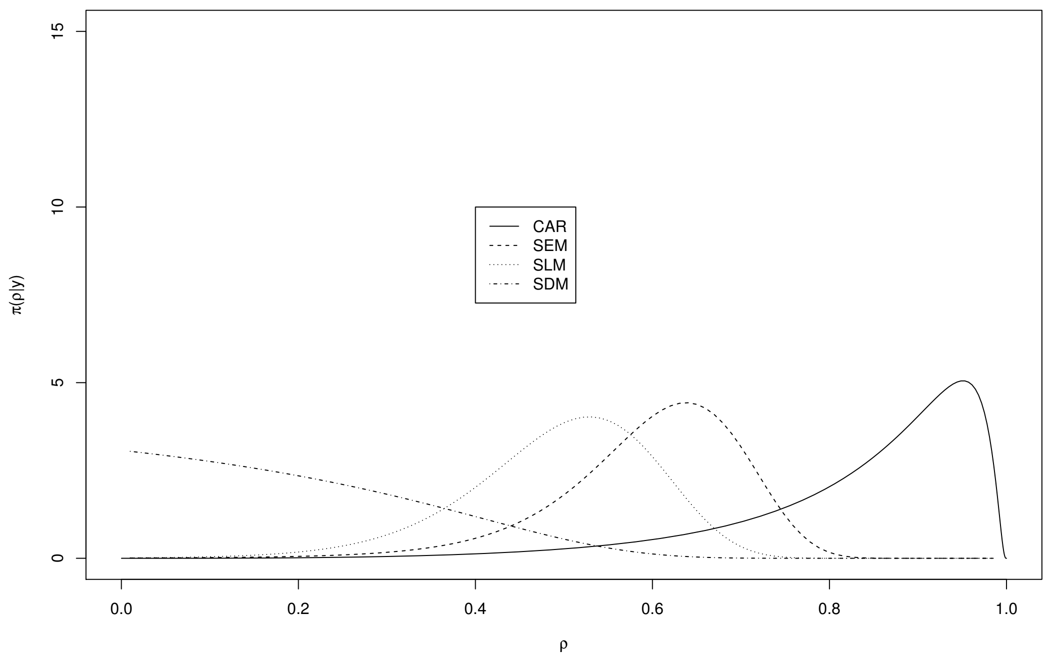

Figure 2: Marginal distributions for the spatial autocorrelation parameter ρ under different models (after-Katrina business data).

Table 14: Total impacts for the SLM model (after-Katrina business data).

	mean	Lower 0.05	Upper 0.95
flood	-0.086	-0.113	-0.060
log(medinc)	0.323	0.138	0.503
sizeempsmall	-0.183	-0.330	-0.050
sizeemplarge	-0.192	-0.526	0.116
sesstatuslow	-0.216	-0.367	-0.074
sesstatushigh	0.037	-0.086	0.162
owntypesole	0.369	0.178	0.586
owntypenational	0.071	-0.289	0.429

Table 15: Indirect impacts for the SLM model (after-Katrina business data).

	mean	Lower 0.05	Upper 0.95
flood	-0.041	-0.058	-0.025
log(medinc)	0.151	0.060	0.256
sizeempsmall	-0.087	-0.175	-0.019
sizeemplarge	-0.093	-0.276	0.054
sesstatuslow	-0.103	-0.194	-0.029
sesstatushigh	0.018	-0.042	0.083
owntypesole	0.175	0.062	0.310
owntypenational	0.034	-0.145	0.219

$$\begin{array}{rcl} y & = & \rho_{\text{Lag}} W y + X\beta + u \\ u & = & \rho_{\text{Err}} M u + \varepsilon \\ \varepsilon & \sim & N(0, \sigma^2 I_n). \end{array} \tag{54}$$

Here, W and M represent two different spatial weight matrices, associated to spatial autocorrelation parameters ρ_{Lag} and ρ_{Err}, respectively. This model is useful when spatial correlation on the response and the residuals are thought to occur at two different spatial scales and they need to be modelled in a different way.

This model can be fitted following the approach described in this paper by conditioning on pairs of values of the spatial autocorrelation parameters ρ_{Lag} and ρ_{Err}. Note that in this case it is necessary to deal with the determinants of the form $|I_n - \rho_{\text{Lag}} W|$ and $|I_n - \rho_{\text{Err}} M|$.

Another interesting model includes spatial autocorrelation on the response at two different spatial levels, so that two spatial correlatation parameters and associated weight matrices are used LeSage and Pace [13, page 151]:

$$\begin{array}{rcl} y & = & \rho_{\text{Lag}} W y + \rho'_{\text{Lag}} V y + X\beta + u \\ u & = & \rho_{\text{Err}} M u + \varepsilon \\ \varepsilon & \sim & N(0, \sigma^2 I_n). \end{array} \tag{55}$$

ρ'_{Lag} is a new spatial autocorrelation parameter and V another weight to model spatial dependence at a different administrative level. Hence, spatial autocorrelation on the response at two different administrative scales can be included in the model.

Now we have three different spatial autocorrelation models. Fitting this model using Bayesian model averaging can be tricky as we need to condition on triplets of values for ρ_{Lag}, ρ'_{Lag} and ρ_{Err}. Furthermore, we will need to have to deal with determinants of the form $|I_n - \rho_{\text{Lag}} W - \rho'_{\text{Lag}} V|$ and $|I_n - \rho_{\text{Err}} M|$.

Finally, LeSage and Pace [13, Chapter 6] discuss model comparison in Spatial Econometrics. They focus on two particular cases: comparing the same structural model with different weight matrices and comparing models with different explanatory variables. As a first approach, the marginal likelihood of the fitted models could be used, so that models with higher values are preferred. In addition to the marginal likelihood, **R-INLA** reports the Deviance Information Criterion [DIC, 24] and Conditional Predictive Ordinate [CPO, 22] for model comparisson and selection.

Bayesian model selection techniques can be used to compare the same model based on different weight matrices LeSage and Pace [13, Section 6.3.1] and compute model probabilities similarly as described in Section 4.1. Namely, say we have m models $\mathcal{M}_1, \ldots, \mathcal{M}_m$ based on m different weight matrices. The associated marginal likelihoods can be written as $\pi(y|\mathcal{M}_j);\ j = 1, \ldots, m$. Hence, following LeSage and Pace [13, Section 6.3.1] and assuming that all m models are equally probable (i.e., the prior probabilities $\pi(\mathcal{M}_j) = 1/m;\ j = 1, \ldots, m$) we can write down the model posterior probabilities as:

$$\pi(\mathcal{M}_i|y) = \frac{\pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i)}{\pi(y)} = \frac{\pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i)}{\sum_{j=1}^{m} \pi(y|\mathcal{M}_j)\pi(\mathcal{M}_j)} = \frac{\pi(y|\mathcal{M}_i)}{\sum_{j=1}^{m} \pi(y|\mathcal{M}_j)}. \quad (56)$$

For comparing models with different (explanatory) variables, LeSage and Pace [13, Section 6.3.2] propose the use of *reversible jump* MCMC [7], which is similar to a Bayesian model averaging method but without the need to ellicitate and fit all possible models. For a small number of variables, posterior model probabilities can be computed as in the case of models with different weight matrices. Unfortunately, implementing this approach with **R-INLA** is not feasible as the number of models grows exponentially with the number of explanatory variables. For example, 15 explanatory variables will produce 32768 different models, which are difficult to evaluate even with INLA. Stepwise selection procedures based on the marginal likelihood or the DIC can be easily implemented using **R-INLA** but this may not lead to the optimal model. This procedure is particularly problematic when there is collinearity among the covariates. This happens, for example, when they have similar spatial patterns.

8. Discussion and final remarks

Bayesian spatial econometric models play an important role in the analysis of data with spatial structure. In this paper we have shown an alternative model fitting approach based on the Integrated Nested Laplace Approximation (INLA). Although it only provides marginal inference, INLA is computationally faster than MCMC and its implementation in the **R-INLA** package for the R programming language can fit a wide range of models. Although the **R-INLA** package cannot fit the spatial econometrics models discussed in this paper we have shown how it is possible to fit conditional models (by fixing one or more parameters in the full model) that can be combined by means of Bayesian Model Averaging to obtain the posterior marginals of the model of interest. In our case, the conditional models arise by fixing the spatial autocorrelation parameter ρ, but this approach can be generalised to other models where more than one parameter is fixed. This involves fitting many different models, but this is not really a computational burden as computations can be parallelized to reduce computational time.

We have compared standard spatial econometrics models and other models based on Generalised Linear Models with spatially correlated random effects. Our simulations show that model fitting using GLMs with random effects provides worse results to those provided by Spatial Econometrics models. This may be due to the different ways in which these models consider spatial dependence.

In the last part of this paper, we have analysed data from examples previously discussed by other authors. In general, our models provide similar results to those obtained in other papers.

The use of INLA can be extended to many other types of econometrics models. Spatio-temporal econometrics models can also be developed with our approach and it would involve only minor modifications to the R code that we have developed. The use of INLA has other advantages that do not require any further work. For example, it is easy to include other types of additive effects on the covariates (such as splines or random effects) in our models. **R-INLA** provides a wide range of priors that can be used in our models as well. If we have missing observations in our dataset, we can easily obtain their predictive distribution.

As future work, we are planning an implementation of the methods described in this paper within the **R-INLA** software as latent models. This will provide an easier interface and access to all the options within this software (prior specifications, tuning options, etc.)

Acknowledgements

Virgilio Gómez-Rubio acknowledges partial support from the Spanish government (MICINN, AGL2011-27747 research project) and FEDER funds.

References

[1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Springer.

[2] Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science 89*(1), 3–25.

[3] Besag, J., J. York, and A. Mollie (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*(1), 1–59.

[4] Cliff, A. D. and J. K. Ord (1981). *Spatial Processes: Models & Applications*. Pion.

[5] Eidsvik, J., A. O. Finley, S. Banerjee, and H. Rue (2012). Approximate bayesian inference for large spatial datasets using predictive process models. *Computational Statistics & Data Analysis 56*(6), 1362 – 1380.

[6] Gómez-Rubio, V., R. S. Bivand, and H. Rue (2014). Spatial models using laplace approximation methods. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, pp. 1401–1417. Springer Berlin Heidelberg.

[7] Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 4*(82), 711–732.

[8] Griffith, D. A. (2004a). Faster maximum likelihood estimation of very large spatial autoregressive models: an extension of the smirnov-anselin result. *Journal of Statistical Computation and Simulation 74*(12), 855–866.

[9] Griffith, D. A. (2004b). A spatial filtering specification for the autologistic model. *Environment and Planning A 36*, 1791–1811.

[10] Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.

[11] Harrison, D. and D. L. Rubinfeld (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management 5*, 81–102.

[12] Hoeting, J., A. R. David Madigan and, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*, 382–401.

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

[14] LeSage, J. P., R. Kelley Pace, N. Lam, R. Campanella, and X. Liu (2011). New orleans business recovery in the aftermath of hurricane katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 174*(4), 1007–1027.

[15] Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society, Series B 73*(4), 423–498.

[16] Mardia, K. V. and A. J. Watkins (1989). On multimodality of the likelihood in the spatial linear model. *Biometrika 76*(2), 289–295.

[17] Nelder, J. and R. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General) 3*(135), 370–384.

[18] Pace, R. K. and O. Gilley (1997). Using the spatial configuration of the data to improve estimation. *Journal of the Real Estate Finance and Economics 14*, 333–340.

[19] R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

[20] Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge.

[21] Ritter, C. and M. A. Tanner (1992). Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. *Journal of the American Statistical Association 87*(419), 861–868.

[22] Roos, M. and L. Held (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *BAYESIAN ANALYSIS 6*(2), 259–278.

[23] Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B 71*(Part 2), 319–392.

[24] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B 64*(4), 583–616.