Article

Estimating Spatial Econometrics Models with Integrated Nested Laplace Approximation

Virgilio Gómez-Rubio [1,*], Roger S. Bivand [2] and Håvard Rue [3]

[1] Department of Mathematics, School of Industrial Engineering, University of Castilla-La Mancha, Albacete, Spain; Virgilio.Gomez@uclm.es
[2] Department of Economics, Norwegian School of Economics, Bergen, Norway.; Roger.Bivand@nhh.no
[3] CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; haavard.rue@kaust.edu.sa
* Correspondence: Virgilio.Gomez@uclm.es; Avda. España s/n, Department of Mathematics, School of Industrial Engineering, University of Castilla-La Mancha, 02071 Albacete

Abstract: The integrated Nested Laplace Approximation [1] provides a fast and effective method for marginal inference on Bayesian hierarchical models. This methodology has been implemented in the **R-INLA** package which permits INLA to be used from within R statistical software. Although INLA is implemented as a general methodology, its use in practice is limited to the models implemented in the **R-INLA** package. Spatial autoregressive models are widely used in spatial econometrics but have until now been missing from the **R-INLA** package. In this paper, we describe the implementation and application of a new class of latent models in INLA made available through **R-INLA**. This new latent class implements a standard spatial lag model. The implementation of this latent model in **R-INLA** also means that all the other features of INLA can be used for model fitting, model selection and inference in spatial econometrics, as will be shown in this paper. Finally, we will illustrate the use of this new latent model and its applications with two datasets based on Gaussian and binary outcomes.

Keywords: Bayesian infernece; INLA; R; spatial econometrics; spatial statistics

Citation: Gómez-Rubio, V.; Bivand, R. S.; Rue, H. Estimating Spatial Econometrics Models with Integrated Nested Laplace Approximation. *Mathematics* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Interest in the Bayesian analysis of regression models with spatial dependence has existed since spatial econometrics came into being in the late 1970s. Hepple [2] and Anselin [3] point to key benefits, such as being able to make exact, finite-sample inferences in models in which only large sample, asymptotic inferences would be feasible, and in the examination of model robustness to specification error [2, p. 180]. Anselin [4, pp. 88–91] extends this discussion, but admits that Bayesian approaches had not at that time been applied often in spatial econometrics. Hepple [5,6] continued to follow up topics within Bayesian estimation, including Bayesian model choice [7]. No software was available until the Spatial Econometrics Library was made available within the Econometrics Toolbox for Matlab, based on LeSage [8,9].[1]

LeSage and Pace [10] provide a summary of spatial econometrics models and their applications. For Bayesian inference, they used Markov Chain Monte Carlo (MCMC) algorithms to estimate the posterior distributions of the model parameters. These techniques give a feasible way of fitting Bayesian models, but can be computationally intensive. In addition, while the Matlab MCMC implementation does provide user access to change prior values from their defaults, the time required to check non-default settings may be considerable.

Bivand *et al.* [11,12] describe how to use the Integrated Nested Laplace Approximation [INLA, 1] to fit some spatial econometrics models. They focus on some models

[1] See http://www.spatial-econometrics.com/.

based on spatial autoregressive specifications on the response and the error terms often used in spatial econometrics. Because of the lack of an implementation of these models within the **R-INLA** software at that time, Bivand *et al.* [11,12] fit many different models conditioning on values of the spatial autocorrelation parameter. These conditional models can be fitted with **R-INLA** and they are later combined using Bayesian model averaging [BMA, 13] to obtain the posterior marginals of the parameters of the desired model. [14] discuss different approaches to fit spatial econometrics with INLA and how to perform multivariate inference on the output. Similarly, [15] exploit Bayesian model averaging to fit spatial econometrics models wiht INLA.

INLA is based on approximating the posterior marginal distributions of the model parameters by means of different Laplace approximations. This provides a numerically fast method to fit models that can be applied to a wide range of research topics. INLA is restricted to models whose latent effects are Gaussian Markov Random Fields, but this class of models includes many models used in practice in a range of disciplines.

In this paper we describe the implementation of a new latent class, that we will call `slm`, within **R-INLA** that facilitates fitting spatial econometrics models. This provides an alternative to fitting some of the models in the Spatial Econometrics Library. In addition, this creates a faster approach for Bayesian inference when only marginal inference on the model parameters is required.

This new approach will make fitting a wide range of spatial econometrics models very easy thought the **R-INLA** package. A flexible specification of these models will allow the inclusion of smooth terms to explore non-linear relationships between variables. The new latent effects for spatial econometrics can be combined with other latent effects to fit more complex models. Furthermore, models will be fitted faster than with traditional MCMC, so a larger number of models can be explored and different model selection techniques can be used.

This paper has the following structure. After providing background descriptions of some spatial econometrics models and the Integrated Nested Laplace Approximation, we introduce the new `slm` latent model in Section 3. A summary on the use of different likelihoods is included in Section 4. The computation of the impacts is laid out in Section 5. Section 6.1 describes some applications on model selection and section 6.2 deals with other issues in Bayesian inference. Examples are included in Section 7, using the well-known Boston house price data set and the Katrina business re-opening data set, and a final discussion is given in Section 8.

2. Background

2.1. Spatial Econometrics Models

In this section we summarise some of the spatial econometrics models that we will use throughout this paper. For a review on spatial econometrics models see Anselin [16]. We will follow the notation used in Bivand *et al.* [11], which is in turn derived from Anselin [4] and LeSage and Pace [10], LeSage and Pace [17].

We will assume that we have a vector y of observations from n different regions. The adjacency structure of these regions is available in a matrix W, which may be defined in different ways. Unless otherwise stated, we will use standard binary matrices to denote adjacency between regions, with standardised rows. This is helpful in offering known bounds for the spatial autocorrelation parameters [see, 18, for details]. Also, we will assume that the p covariates available are in a design matrix X, which will be used to construct regression models. Pace *et al.* [19] have pointed out challenges involved in estimating models of this kind that we intend to address in further research.

The first model that we will describe is the Spatial Error Model (SEM) which is based on a spatial autoregressive error term:

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

Here, ρ_{Err} is the spatial autocorrelation parameter associated to the error term. This measures how strong spatial dependence is. β is the vector of coefficients of the covariates in the model. The error term e is supposed to follow a multivariate Normal distribution with zero mean and diagonal variance-covariance matrix $\sigma^2 I_n$. σ^2 is a global variance parameter while I_n is the identity matrix of dimension $n \times n$.

Alternatively, we can consider an autoregressive model on the response (Spatial Lag Model, SLM):

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

ρ_{Lag} is now the spatial autocorrelation parameter associated to the autocorrelated term on the response.

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

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

γ is a vector of coefficients for the spatially lagged covariates, shown as matrix WX.

A variation of this model is the Spatial Durbin Error Model (SDEM), in which the error is autoregressive:

$$y = X\beta + WX\gamma + u;\ u = \rho_{\text{Err}} M u + e;\ e \sim MVN(0, \sigma^2 I_n). \quad (4)$$

Here M is an adjacency matrix for the error term that may be different from W.

All these models can be rewritten so that the response y only appears on the left hand side. The SEM model can also be written as

$$y = X\beta + (I_n - \rho_{\text{Err}} W)^{-1} e; e \sim MVN(0, \sigma^2 I_n); \quad (5)$$

the SLM model is equivalent to

$$y = (I_n - \rho_{\text{Lag}} W)^{-1} (X\beta + e);\ e \sim MVN(0, \sigma^2 I_n); \quad (6)$$

the SDM model is

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

with $X^* = [X, WX]$, the new matrix of covariates with the original and the lagged covariates and $\beta' = [\beta, \gamma]$, the associated vector of coefficients. Finally, the SDEM can be written as

$$y = X^* \beta' + (I_n - \rho_{\text{Err}} M)^{-1} e;\ e \sim MVN(0, \sigma^2 I_n). \quad (8)$$

For completeness, we will include a simplified model without spatial autocorrelation parameters and lagged variables (Spatially Lagged X model, SLX):

$$y = X^* \beta' + e;\ e \sim MVN(0, \sigma^2 I_n), \quad (9)$$

These are a set of standard models in spatial econometrics, focussing on three key issues: spatially autocorrelated errors, spatially autocorrelated responses and spatially lagged covariates. More complex models can be built from these three standard models; the main difference is that those models incorporate more than one spatial autocorrelation parameter.

2.2. The Integrated Nested Laplace Approximation

Bayesian inference on hierarchical models has often relied on the use of computational methods among which Markov Chain Monte Carlo is the most widely used. In principle, MCMC has the advantage of being able to handle a large number of models, but it has it drawbacks, such as slow convergence of the Markov chains and the difficulty of obtaining sampling distributions for complex models.

Rue *et al.* [1] have developed an approximate method to estimate the marginal distributions of the parameters in a Bayesian model. In particular, they focus on the family of Latent Gaussian Markov Random Fields models. We describe here how this new methodology has been developed, but we refer the reader to the original paper for details.

First of all, a vector of n observed values $\mathbf{y} = (y_1, \ldots, y_n)$ are assumed to be distributed according to one of the distributions in the exponential family, with mean μ_i. Observed covariates and a linear predictor on them (possibly plus random effects) may be linked to the mean μ_i by using an appropriate transformation (i.e., a link function). Hence, this linear predictor η_i may be made of a fixed term on the covariates plus random effects and other non-linear terms.

The distribution of $\mathbf{y}$ will depend on a number of hyperparameters $\boldsymbol{\theta}_1$. The vector $\mathbf{x}$ of latent effects forms a Gaussian Markov Random Field with precision matrix $Q(\boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_2$ is a vector of hyperparameters. The hyperparameters can be represented in a unique vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. It should be noted that observations $\mathbf{y}$ are independent given the values of the latent effects $\mathbf{x}$ and the hyperparameters $\boldsymbol{\theta}$. This can be written as

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i \in \mathcal{I}} \pi(y_i | x_i, \boldsymbol{\theta}) \tag{10}$$

Here, x_i represents the linear predictor η_i and $\mathcal{I}$ is a vector of indices over $1, \ldots, n$. If there are missing values in $\mathbf{y}$ these indices will not be included in $\mathcal{I}$.

The posterior distribution of the latent effects $\mathbf{x}$ and the vector of hyperparameters $\boldsymbol{\theta}$ can be written as

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i | x_i, \boldsymbol{\theta}) = \tag{11}$$

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

INLA will not try to estimate the joint distribution $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ but the marginal distribution of single latent effects and hyperparameters, i.e., $\pi(x_j|\mathbf{y})$ and $\pi(\theta_k|\mathbf{y})$. Indices j and k will move in a different range depending on the number of latent variables and hyperparameters.

INLA will first compute an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$, $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, that will be used later to compute an approximation to $\pi(x_j|\mathbf{y})$. This can be done because

$$\pi(x_j|\mathbf{y}) = \int \pi(x_j|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}. \tag{12}$$

Hence, an approximation can be developed as follows:

$$\tilde{\pi}(x_j|\mathbf{y}) = \sum_g \tilde{\pi}(x_j|\boldsymbol{\theta}_g, \mathbf{y}) \times \tilde{\pi}(\boldsymbol{\theta}_g|\mathbf{y}) \times \Delta_g, \tag{13}$$

Here, $\boldsymbol{\theta}_g$ are values of the ensemble of hyperparameters in a grid, with associated weights Δ_g. $\tilde{\pi}(x_j|\boldsymbol{\theta}_g, \mathbf{y})$ is an approximation to $\pi(x_j|\boldsymbol{\theta}_g, \mathbf{y})$ and this is thoroughly addressed in Rue *et al.* [1]. They comment on the use of a Gaussian approximation and others based on repeated Laplace Approximations and explore the error of the approximation.

This methodology is implemented in the **R-INLA** package. It allows for an easy access to many different types of likelihoods, latent models and priors for model fitting. However, this list is by no means exhaustive and there are many latent effects that have not been implemented yet. This is the reason why we describe a newly implemented `slm` latent effect that has many applications in spatial econometrics.

3. The `slm` latent model in R-INLA

Although the INLA methodology covers a wide range of models, latent models need to be implemented in compiled code in the INLA software to be able to fit the models described earlier in this paper. Hence, the newly implemented `slm` latent model fills the gap for spatial econometrics models. This new latent model implements the following expression as a random effect that can be included in the linear predictor:

$$\mathbf{x} = (I_n - \rho W)^{-1}(X\beta + \varepsilon) \tag{14}$$

Here, $\mathbf{x}$ is a vector of n random effects, I_n is the identity matrix of dimension $n \times n$, ρ is a spatial autocorrelation parameter (that we will discuss later), W is a $n \times n$ weight matrix, X a matrix of covariates with coefficients β and ε is a vector of independent Gaussian errors with zero mean and precision τI_n.

In this latent model, we need to assign prior distributions to the vector of coefficients β, spatial autocorrelation parameter ρ and precision of the error term τ. By default, β takes a multivariate Gaussian distribution with zero mean and precision matrix Q (which must be specified); $\text{logit}(\rho)$ takes a Gaussian prior with zero mean and precision 10; and, $\log(\tau)$ takes a log-gamma prior with parameters 1 and $5 \cdot 10^{-5}$. Other priors can be assigned to these hyperparameters following standard **R-INLA** procedures.

Note that, as described in Section 2.1, spatial econometrics models can be derived from this implementation. In particular, the SEM model is a particular case of equation (14) with $\beta = 0$. The SLM model can be fitted with no modification and the SDM model can be implemented using a matrix of covariates made of the original covariates plus the lagged covariates.

The SDEM model simply takes two terms, a standard linear term on the covariates (and lagged covariates), plus a `slm` effect with $\beta = 0$. Finally, the SLX model can be fitted using a standard linear regression on the covariates and lagged covariates and typical i.i.d. random effects.

3.1. Implementation

We will describe here the implementation of the new `slm` latent class. For a Gaussian response (and similarly for non-Gaussian likelihoods) the model can be written as

$$y = (I - \rho W)^{-1}(X\beta + \varepsilon)$$

It can be rewritten using $\mathbf{x} = (I - \rho W)^{-1}(X\beta + \varepsilon)$, so that with observations y, then we have

$$y = \mathbf{x} + e$$

where e is a tiny error that is introduced to fit the model. This is the error present in a Gaussian distribution and will not appear if another likelihood is used.

By re-writing the `slm` as $\mathbf{x}$ in this way, we define it so that it suits the `f()`-component in the **R-INLA** framework. Given this, we note that the `slm` model is a Markov model with a sparse precision matrix, and so conforms to the INLA framework. We provide a detailed proof in B and show here the main results.

The mean and precision of $(\mathbf{x}, \beta)$ given the hyperparameters τ and ρ are given by

$$E[\mathbf{x}, \beta | \tau, \rho] = 0 \tag{15}$$

$$Prec[\mathbf{x}, \beta | \tau, \rho] = \begin{bmatrix} \tau(I_n - \rho W')(I_n - \rho W) & -\tau(I_n - \rho W')X \\ -\tau X'(I_n - \rho W) & Q + \tau X'X \end{bmatrix} \tag{16}$$

Note that this precision matrix is highly sparse and symmetric. Efficient computation using this new latent effect can be carried out using the GMRF library, as described in Rue and Held [20].

The full model can then be derived conditioning on different parameters. Hence, the joint distribution of $\mathbf{x}$, β, τ and ρ can be written as

$$\pi(\mathbf{x}, \beta, \tau, \rho) = \pi(\mathbf{x}, \beta|\tau, \rho)\pi(\tau, \rho) = \pi(\mathbf{x}, \beta|\tau, \rho)\pi(\tau)\pi(\rho)$$

$\pi(\mathbf{x}|\beta, \rho)$ is a Gaussian distribution with mean and precision shown in equations (15) and (16), respectively. The prior distribution of β is Gaussian with zero mean and known precision matrix Q. This matrix is often taken diagonal to assume that the coefficients are independent a priori. Also, it may be worth rescaling the covariates in X to avoid numerical problems. Including lagged covariates may lead to further numerical instability as they may be highly correlated with the original covariates.

It is internally assumed that ρ is between 0 and 1 so that a Gaussian prior is assigned to $\log(\rho/(1-\rho))$. When computing $I_n - \rho W$, ρ is re-scaled to be in a range of appropriate values. See details in the description of the R interface in A.

The simulation study provided in one of the Appendices shows that the implementation works well in practice. A more thorough simulation study could be provided but this is out of the scope of this paper. Furhtermore, the results presented in the examples in Section 7 also support that the new `slm` latent effect provides good estimates of the model parameters.

4. Using different likelihoods

4.1. Binary response

The models described in Section 2.1 assume a Gaussian response but other distributions can be used for the response. LeSage *et al.* [21] consider a binary outcome when studying the probability of re-opening a business in New Orleans in the aftermath of Hurricane Katrina. Binary outcome y_i is modelled using a latent Gaussian variable y_i^* as follows:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases} . \quad (17)$$

y_i^* is the net profit, so that if it is equal or higher than zero the business will re-open. y_i^* is assumed to be a Gaussian variable that can be modelled using the spatial econometrics models described in Section 2.1. Note that in this case the variance of the error term is set to 1 to avoid identifiability problems [11,21].

Several authors have assessed different methods for the estimation of spatial probit models. Billé [22] has compared the methods of Maximum Likelihood (ML) and Generalised Method of Moments using a Monte Carlo study and propose alternatives to avoid the inversion of $|I - \rho W|$ when fitting the models. Calabrese and Elkink [23] compare a larger number of estimation methods for spatial probit models, including MCMC algorithms, to estimate the spatial autocorrelation parameter and provide a study on the predictive quality of each method.

Instead of using a "broken-stick" function such as the one shown in equation (17), **R-INLA** relies on standard logit and probit links, among others. In our examples, the spatial probit is based on using a (continuous) probit link function instead of the one shown in equation (17). Hence, differences in some results can be expected.

4.2. Other likelihoods

R-INLA provides a number of likelihoods that can be used when defining a model, so that an `slm` latent effect is included in the linear predictor. However, attention should be paid so that the resulting model makes sense. A particular problem of interest is whether all parameters in the model are identifiable. For example, if the spatial probit model is used, τ must be set to one so that β can be properly estimated [10]. Using other highly parameterised likelihoods (such as, for example, zero-inflated models) obliges the analyst to pay attention to details to ensure that output makes sense.

4.3. Additional effects

R-INLA makes it possible to include additional effects in the linear predictor. All models presented so far assume a spatial structure on the error terms. Like Besag *et al.* [24], it is possible to consider a model in which there are two different random effects: one spatial with an autocorrelated structured defined by the `slm` latent class plus unstructured random effects. For example, the SLM model can be extended as follows:

$$y = (I_n - \rho_{\text{Lag}} W)^{-1}(X\beta + \epsilon) + u;\ u \sim MVN(0, \sigma^2 I_n) \quad (18)$$

Note that the random effect u can have different structures.

Furthermore, other effects than linear can be explored for the covariates, as **R-INLA** includes different types of smoothers, such as first and second order random walks.

5. Computation of the Impacts

Impacts are related to how changes in the covariates in one area will affect the response in other areas, and there are different types of *impacts* to measure these effects. As LeSage and Pace [10, Section 2.7] explain, impacts appear because a change of the value of a covariate in a region will affect not only the region itself (direct impact) but also other regions indirectly (indirect impact).

For the linear models presented in Section 2.1, impacts per covariate r can be defined as

$$\frac{\partial E[y_i \mid x_r]}{\partial x_{jr}} \quad i, j = 1, \ldots, n;\ r = 1 \ldots, p \quad (19)$$

with x_r the r-th covariate. This will measure the change in the response observed in area i when covariate r is changed in area j. For the spatial probit models, the impacts are defined as [21]:

$$\frac{\partial Pr(y_i = 1)}{\partial x_{jr}} \quad i, j = 1, \ldots, n;\ r = 1 \ldots, p \quad (20)$$

In both cases, the impacts will produce a $n \times n$ matrix of impacts $S_r(W)$ for each covariate. The values on the diagonal of this matrix are called *direct impacts*, as they measure the change in the response when the covariate is changed at the same area (i.e., the value of covariate r in area i is changed). In order to give an overall measure of the direct impacts, its average is often computed and it is called *average direct impact*.

Similarly, *indirect impacts* are defined as the off-diagonal elements of $S_r(W)$, and they measure the change in the response in one area when changes in covariate r happen at any other area. A global measure of the indirect impacts is the sum of all off-diagonal elements divided by n, which is called the *average indirect impact*.

Finally, the *average total impact* is defined as the sum of direct and indirect impacts. This gives an overall measure of how the response is affected when changes occur in a covariate at any area.

Bivand *et al.* [11] summarise the form of the impacts for different models and provide some ideas on how to compute the different average impacts with **R-INLA** and BMA. For a Gaussian response, the impacts matrix for the SEM model is simply a diagonal matrix with coefficient β_r in it, i.e., $S_r(W) = I_n \beta_r$. For the SDM model, the impacts matrix is

$$S_r(W) = (I_n - \rho_{\text{Lag}} W)^{-1}(I_n \beta_r + W\gamma_r);\ r = 1, \ldots, p. \quad (21)$$

The impact matrix for the SLM model is the same as in equation (21) with $\gamma_r = 0,\ r = 1, \ldots, v$, i.e., the coefficients of the lag covariates are not considered.

In addition, the impacts matrix for the SDEM model is

$$S_r(W) = (I_n \beta_r + W\gamma_r);\ r = 1, \ldots, p \quad (22)$$

Finally, the SLX model shares the same impacts matrix as the SDEM model; in both cases, the average impacts are the coefficients β_r (direct) and γ_r (indirect), for which inferences are readily available.

In the case of the spatial probit, the impacts matrices are similar but they need to be premultiplied by a diagonal matrix $D(f(\eta))$, which is a $n \times n$ diagonal matrix with entries $f(\eta_i), i = 1, \ldots, n$, where $f(\eta_i)$ is the standard Gaussian distribution evaluated at the expected value of the linear predictor of observation i. For example, for the spatial probit SDM model this is:

$$S_r(W) = D(f(\eta))(I_n - \rho_{\text{Lag}} W)^{-1}(I_n \beta_r + W\gamma_r);\ r = 1, \ldots, v, \tag{23}$$

where η is defined as

$$\eta = (I_n - \rho_{\text{Lag}} W)^{-1}(X\beta_r + WX\gamma_r);\ r = 1, \ldots, v. \tag{24}$$

The impacts matrix for other spatial probit models can be derived in a similar way.

5.1. Approximation of the impacts

Average direct, indirect and total impacts can be computed by summing over the required elements in the impacts matrix (and dividing by n). In a few simple cases, such as the SEM, SDEM and SLX models, the impacts can be computed with **R-INLA** as the impacts are a linear combination of the covariate coefficients. In general, the impacts cannot be computed directly with **R-INLA** as they are a function on several parameters and INLA only provides marginal inference.

From a general point of view, the average impacts can be regarded as the computation of a functions that involves two or three parameters in the model. Hence, multivariate posterior inference is required to obtain estimates of the impacts. We will try to approximate the posterior marginal of the average impacts the best we can by sampling from the approximate joint posterior (see below).

Let us consider the Gaussian case first. For the SEM, the average direct impact for covariate r is simply the posterior marginal of coefficient β_r. So this is a trivial case and inference is exact. Average indirect impacts are equal to zero, which makes the average total impact equal to the average direct impact, i.e., β_r.

For the SDM model, the average total impact is

$$\frac{1}{1 - \rho_{\text{Lag}}}(\beta_r + \gamma_r) \tag{25}$$

Note how this is the product of two terms, one on ρ_{Lag} and the other one on $\beta_r + \gamma_r$. In order to estimate the posterior distribution the average total impact, samples from approximate joint posterior of the parameters invovled can be drawn from the fitted INLA model using the `inla.posterior.sample` function [25, Chapter 2].

Regarding the average direct impact for the SDM model, this is

$$n^{-1} tr\Big((I_n - \rho_{\text{Lag}} W)^{-1}\Big)\beta_r + n^{-1} tr\Big((I_n - \rho_{\text{Lag}} W)^{-1} W\Big)\gamma_r \tag{26}$$

Again, this expression is a non-linear term that involves several parameters and its posterior distribution can be approximated by sampling from the approximate posterior distribution obtained with INLA.

For the SDEM and SLX models, the distribution of the average total impact is the marginal distribution of $\beta_r + \gamma_r$, whilst the associated direct impact is given by

$$n^{-1} tr\Big(I_n\Big)\beta_r + n^{-1} tr\Big(W\Big)\gamma_r = \beta_r \tag{27}$$

Hence, inference on the impacts is exact for the SDEM and SLX models.

In the case of the spatial probit, the average total impacts are as before but multiplied by

$$\sum_{i=1}^{n} \frac{f(\eta_i)}{n} \tag{28}$$

The average direct impact is the trace of $S_r(W)$, which now takes a more complex form as it involves $D(f(\eta))$, divided by n. For example, for the SDM model it is

$$\begin{aligned} n^{-1}tr[D(f(\eta))(I_n - \rho_{\text{Lag}}W)^{-1}(\beta_r + W\gamma_r)] = \\ n^{-1}tr[D(f(\eta))(I_n - \rho_{\text{Lag}}W)^{-1}]\beta_r + \\ +n^{-1}tr[D(f(\eta))(I_n - \rho_{\text{Lag}}W)^{-1}W]\gamma_r) \end{aligned} \tag{29}$$

The posterior distribution of the impacts will be approximated using sampling from INLA, as stated above.

6. Further topics

6.1. Some applications in Spatial Econometrics

LeSage and Pace [10] not only describe how to fit Bayesian Spatial Econometrics models using MCMC but also discuss how to take advantage of the Bayesian approach to tackle a number of other issues. In this section we aim at discussing other applications when dealing with spatial econometrics models.

6.1.1. Model selection

R-INLA reports the marginal likelihood of the fitted model $\mathcal{M}$, i.e., $\pi(y|\mathcal{M})$, which can be used for model selection, as described in LeSage and Pace [10, Section 6.3] and Bivand *et al.* [11]. For example, if we have a set of m fitted models $\{\mathcal{M}_i\}_{i=1}^m$ with marginal likelihoods $\{\pi(y|\mathcal{M}_i)\}_{i=1}^m$, we may select the model with the highest marginal likelihood as the "best" model.

Following a fully Bayesian approach, we could compute the posterior probability of each model taking a set of prior model probabilities $\{\pi(\mathcal{M}_i)\}_{i=1}^m$ and combining them with the marginal likelihoods using Bayes' rule:

$$\pi(\mathcal{M}_i|y) = \frac{\pi(y|\mathcal{M}_i)\pi(\mathcal{M}_i)}{\sum_{j=1}^m \pi(y|\mathcal{M}_j)\pi(\mathcal{M}_j)} \quad i = 1, \ldots, m \tag{30}$$

If all models are thought to be equally likely a priori then the priors are taken as $\pi(\mathcal{M}_i) = 1/m$, so that the posterior probabilities are obtained by re-scaling the marginal likelihoods to sum up to one:

$$\pi(\mathcal{M}_i|y) = \frac{\pi(y|\mathcal{M}_i)}{\sum_{j=1}^m \pi(y|\mathcal{M}_j)} \quad i = 1, \ldots, m \tag{31}$$

This model selection approach can be applied to models with very different structures. They could be models with different spatial structures, different latent effects or based on different sets of covariates. In Section 7 we show an example based on comparing models with different spatial structures in the second example on the Katrina business data.

In addition, **R-INLA** implements a number of criteria for model selection, such as the Deviance Information Criterion [26, DIC] and Conditional Predictive Ordinate [27, CPO]. These criteria can be used to compare different models and perform model selection as well.

6.1.2. Variable selection

As a particular application of model selection we will discus here how to deal with variable selection. In this case, models differ in the covariates that are included as fixed effects. The number of possible models that appear is usually very large. For example, 20 covariates will produce 2^{20} possible models, i.e., more than 1 million models to be fitted. As stated before, posterior probabilities for each model can be computed using the marginal likelihood as in equation (31). In principle, given that **R-INLA** fits models very quickly and that a large number of models can be fitted in parallel on a cluster of computers, it would be feasible to fit all possible models.

As an alternative approach, stepwise regression can be performed based on any of the model selection criteria available. In particular, the DIC provides a feasible way of performing variable selection. This can be included in a step-wise variable selection procedure which will not explore all possible models but that can lead to a sub-optimal model.

6.1.3. Model averaging

Sometimes, an averaged model may be obtained from other fitted models. We have already pointed out how Bivand *et al.* [11] use Bayesian model averaging to fit spatial econometrics models using other models with simpler random effects. [15] describe the use of Bayesian model averaging with INLA and how to fit a spatial econometrics models with two spatial autoregression parameters. This approach can be employed to fit highly parameterized models with INLA.

However, a BMA approach can also be used to combine different models for other purposes. For example, when the adjacency matrix is unknown we may fit different models using slightly different adjacency matrices. LeSage and Pace [10] discuss BMA in the context of spatial econometrics. In Section 7 we have considered this in the second example on the Katrina business data where different spatial structures are considered using a nearest neighbour algorithm.

6.2. Other issues in Bayesian inference

So far, we have made a review of some existing and widely used Spatial Econometrics models, and how these models can be fitted using INLA and its associated software **R-INLA**. Now, we will focus on other general problems that can be tackled using this new approach.

6.2.1. Linear combinations and linear constraints on the parameters

R-INLA allows the computation of posterior marginals of linear combinations on the latent effects. This can be very useful to compute some derived quantities from the fitted models, such as some of the impacts described in Section 5.

Furthermore, **R-INLA** allows the user to define linear constraints on any of the latent parameters and other quantities, including the linear predictor. This is useful, for example, to produce benchmarked estimates, i.e., model-based estimates obtained at an aggregation level that must match a particular value at a different aggregation level.

6.2.2. Prediction of missing values in the response

Missing values often appear in spatial econometrics because actual data have not been gathered for some regions or the respondent was not available at the time of the interview. Sometimes, missing data appear because of the way surveys are designed as the sample is taken to be representative of the whole population under study and many small areas may not be sampled at all. Missing values may also appear in the covariates, but we will only consider here the case of missing values in the response.

With **R-INLA**, a posterior marginal distribution will be obtained and inference and predictions on the missing responses can be made from this. Note that this is a prediction only and that uncertainty about the missing values will not influence the parameter estimates.

The case of missing values in the covariates is more complex. First of all, we will need to define a reasonable imputation model and, secondly, the missing values and the parameters in the new imputation model will be treated as hyperparameters in our approach, increasing the number of hyperparameters and making a computational solution infeasible.

6.2.3. Choice of the priors

LeSage and Pace [10] briefly discuss the choice of different priors for the parameters in the model, and they stress the importance of having vague priors. For the spatial autocorrelation parameter they propose a uniform distribution in the range of this parameter. A Normal prior with zero mean and large variance is used for β. A Gamma prior with small mean and large variance is proposed for the variance σ^2. However wise these choices may seem, it is not clear how these priors will impact on the results.

Because of the way different models in the Spatial Econometrics Toolbox are implemented, it is difficult to assess the impact of the priors, as using different priors will require rethinking how the MCMC sampling is done. New conditional distributions need to be worked out and implemented.

R-INLA provides a simple interface with some predefined priors that can be easily used. Other priors can be defined by using a convenient language and plugged into the **R-INLA** software. Hence, it is easier and faster to assess the impact of different priors. See Chapter 5 in [25] for a general discussion on the use of priors with **R-INLA**.

For example, Gelman [28] has suggested that Gamma priors for the variances were not adequate for the variance parameter in Gaussian models as they were too informative. Instead, they have proposed the use of a half-Cauchy distribution. A model with this prior can be easily implemented by defining the half-Cauchy prior and passing it to **R-INLA**.

7. Examples

7.1. Boston housing data

Harrison and Rubinfeld [29] study the median value of owner-occupied houses in the Boston area using 13 covariates as well. Note that the median value has been censored at $50,000 and that we omit tracts that are censored, leaving 490 observations [30]. The spatial adjacency that we will consider is for census tract contiguities.

Bivand *et al.* [11] use INLA and Bayesian model averaging to fit SEM, SLM and SDM models to this dataset (but including all 506 tracts and using a different representation of adjacency). Here we will use the new `slm` latent model for **R-INLA** to fit the same models. In principle, we should obtain similar results and we will benefit from all the other built-in features in **R-INLA** (such as, summary statistics, model selection criteria, prediction, etc.).

First of all, we have fitted the five models described in Section 2.1. Point estimates of the fixed effects are summarised in Table 1. In addition, the posterior marginal of the spatial autocorrelation parameters have been displayed in Figure 1, including posterior means. Note that these values are not the ones reported in the **R-INLA** output and that we have re-scaled them as explained in Section 3. In this case, the range used for the spatial autocorrelation parameter is $(-1, 1)$. This will make the summary statistics for ρ directly comparable to those reported in Bivand *et al.* [11].

In general, all our results match theirs as expected. However, the new `slm` latent effects makes fitting these models with **R-INLA** simpler. Finally, Figure 2 shows a map of the values of the `slm` latent effects for the SEM and SDEM models.

Note that in order to fit the model we have set the variance of the Gaussian likelihood to a fixed and tiny value (e^{-15}) because this error term does not appear in the different spatial econometrics model fit. A side effect of this is that the DIC will be the same for all models (and it cannot be used for model comparison) and that the fitted values will also have a tiny variance. The fitted values could be computed in the right way by adding extra observations with missing values (i.e., `NA`) in the response; this obsevations will not be used for model fitting and the fitted values will now account for the required uncertainty.

	SEM	SLM	SDM	SDMlag	SDEM	SDEMlag	SLX	SLXlag
(Intercept)	3.542	2.168	1.946		4.364		5.122	
CRIM	-0.007	-0.007	-0.006	-0.003	-0.007	-0.006	-0.007	-0.015
ZN	0.000	0.000	0.000	-0.000	0.000	-0.001	0.000	0.000
INDUS	0.001	0.002	-0.000	0.000	0.001	0.001	-0.002	0.002
CHAS1	-0.047	-0.002	-0.062	0.110	-0.047	0.133	-0.065	0.194
I(NOX^2)	-0.149	-0.232	0.011	-0.404	-0.050	-0.585	0.081	-1.056
I(RM^2)	0.010	0.008	0.010	-0.009	0.010	-0.002	0.009	-0.007
AGE	-0.001	-0.000	-0.001	0.002	-0.001	0.001	-0.001	0.002
log(DIS)	-0.033	-0.140	-0.040	-0.064	-0.082	-0.076	-0.038	-0.221
log(RAD)	0.059	0.062	0.052	-0.004	0.059	0.014	0.043	0.078
TAX	-0.001	-0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
PTRATIO	-0.018	-0.013	-0.013	-0.002	-0.016	-0.016	-0.012	-0.027
B	0.001	0.000	0.001	-0.001	0.000	-0.000	0.001	-0.001
log(LSTAT)	-0.226	-0.218	-0.215	0.054	-0.233	-0.112	-0.234	-0.180

Table 1: Posterior means of the fixed effects coefficients, Boston housing data.

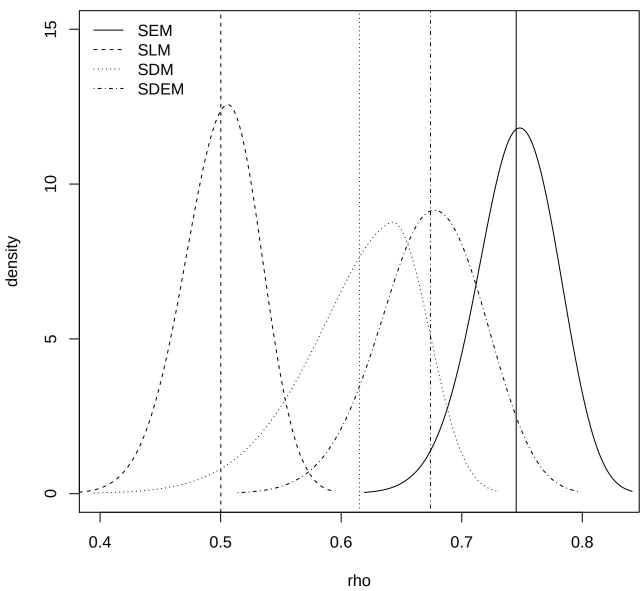

Figure 1. Posterior marginal of the spatial autocorrelation parameters with posterior means (vertical lines), Boston housing data.

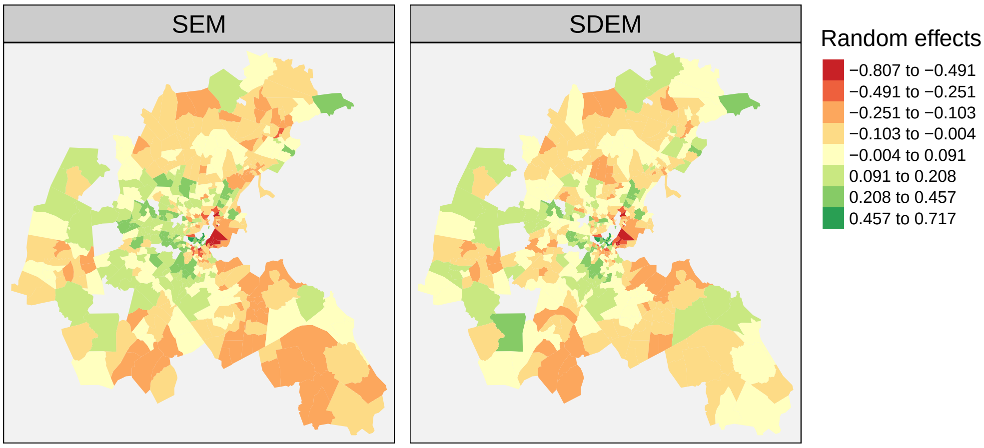

Figure 2. Posterior means of the `slm` latent effects for the SEM and SDEM models, Boston housing data.

7.1.1. Smoothing covariates

So far we have considered the squared values of NOX in our models but the relationship between this covariate and the response may take other forms. In Section 4.3 we have discussed how **R-INLA** implements some latent effects to smooth covariates [see, for example, 25, Chapter 9]. One of them is the second order random walk that we have used here to smooth the values of nitric oxides concentration (parts per 10 million), which is covariate NOX. This smoother needs to be included additively in all the other effects, so it is only readily available for the SEM, SDEM and SLX models. This latent model has only an hyperparameter, which is its precision. In order to avoid overfitting and force the random walk to produce a smooth function, a Gamma prior with mean 2000 and variance 10 has been used for the precision in all models, so that the level of smoothing can be compared. The results are shown in Figure 3.

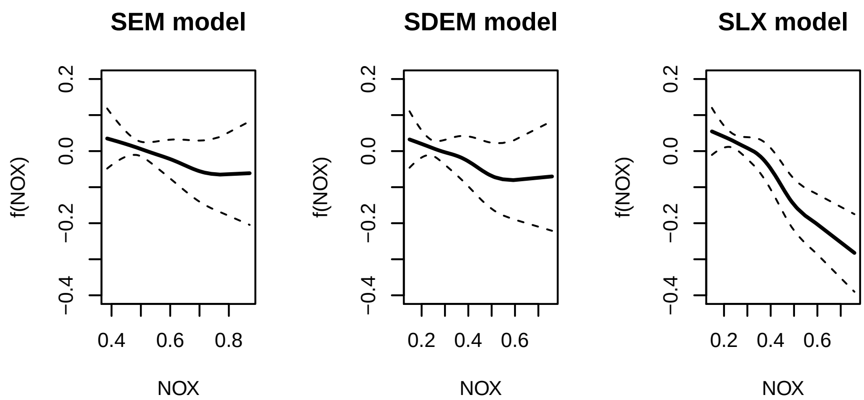

Figure 3. Smoothed effects of nitric oxides concentration (NOX) for three spatial econometrics models with 95% credible intervals.

SEM and SDEM show a similar linear effect, which is consistent with the findings in Harrison and Rubinfeld [29]. The SLX model seems to provide similar estimates of this effect but with considerably narrower credible intervals. As no spatial random effects are included in this model, we believe that the smoother on NOX is picking up residual spatial effects.

7.1.2. Impacts

Average direct, indirect and total impacts have been computed for the models fitted to the Boston housing data set. In addition, for the SLM and SDM models we have also fitted the models using maximum likelihood and computed their impacts for purposes of comparison. Average direct impacts, average total impacts and average indirect impacts are provided as supplementary materials. Impacts are very similar between ML and Bayesian estimation for the models where we have computed them using functioni in the **spdep** package.

Inference on the different impacts is based on their respective posterior marginal distributions provided by **R-INLA**. In addition to the posterior means other statistics can be obtained, such as standard deviation, quantiles and credible intervals. These posterior marginal distributions can also be compared to assess how different models produce different impacts. Figure 4 shows the total impacts for NOX-squared under five different models. Given that for models SLM and SDM we were using an approximation we have included, in a thicker line, the distribution of the total impacts obtained with the Matlab code in the Spatial Econometrics Toolbox for these two models. The results clearly show that our approximation is very close to the results based on MCMC. Furthermore, we have checked that similar accuracy is obtained for all the covariates included in the model.

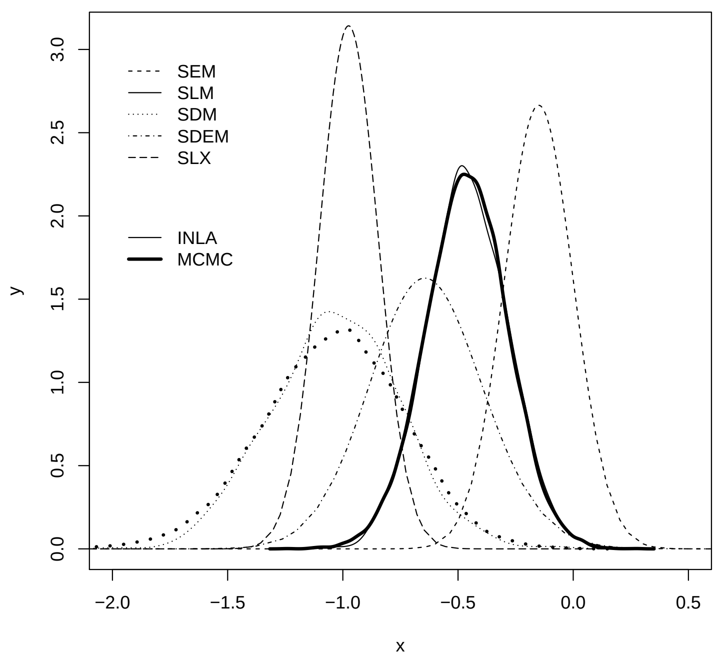

Figure 4. Total impacts for NOX-squared obtained with different models, Boston housing data

7.1.3. Prediction of missing values

As stated in Section 6.2.2, INLA and **R-INLA** can provide predictions of missing values in the response. We will use this feature to provide predictions of the 16 tracts with censored observations of the median values, treated as missing values. This will allow us

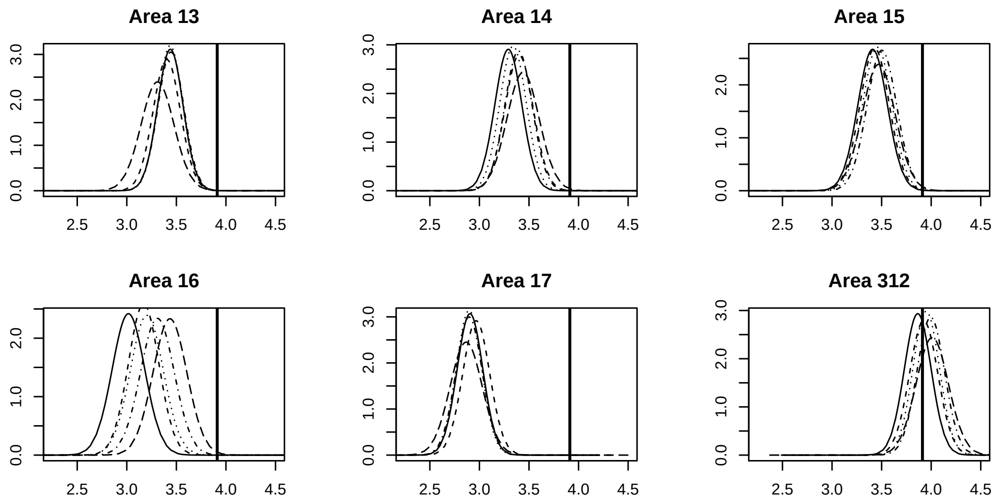

Figure 5. Marginal distributions of the median housing values for 6 of the 16 areas with censored observations, Boston housing data with full adjacency matrix. The legend is as in Figure 4. Vertical lines shows where the censoring cuts in.

to use the complete adjacency structure and to borrow information from neighbouring areas to provide better predictions.

With **R-INLA**, this is as simple as setting the censored values in the response to `NA`. We will obtain a predictive distribution for the missing values so that inference can be made from them. We have represented the five marginal distributions obtained with the models in Figure 5 for 6 selected areas. The vertical line shows where the censoring cuts in.

Areas 13 to 17 seem to have predicted values well below the cut-off point. These ares are located in the city center, where house prices are likely to be higher than average and that is why our model does not predict well there. Furthermore, predicted values the remaining 11 areas with missing values have a similar behaviour as in Area 312 (included in Figure 5), that is, the cut-off point is close to the median of the predicted values.

Furthermore, in the supplementary materials we show the posterior means of the `slm` latent effects for the SEM and SDEM models in the same way as in Figure 2 for the incomplete data set. The main difference is that the new maps include predictions for the areas with the censored observations but the estimates in the common tracts are similar.

7.2. Business opening after Katrina

LeSage *et al.* [21] study the probabilities of reopening a business in New Orleans in the aftermath of hurricane Katrina. They have used a spatial probit, as the one described in equation (17). Here we reproduce the analysis with a continuous link function (i.e., a probit function) and the new `slm` latent model, similarly as in Bivand *et al.* [11].

7.2.1. Standard models

Table 2 shows point estimates (posterior means) of the fixed effects for the five of the models discussed in Section 2.1.

Furthermore, Figure 6 shows the marginal distribution of the spatial autocorrelation parameters of four different spatial econometrics models for INLA and MCMC. For the INLA models, the values of the spatial autocorrelation parameters have been properly re-scaled to fit in the correct range and not constrained to the (0,1) interval. In this example ρ could be betwee from -3.276 to 1, but we have only considered it to be in the $(-1, 1)$ so that a fair comparisson with MCMC can be done. INLA estimates for both fixed and spatial autocorrelation estimates are very similar to those reported in Bivand *et al.* [11]. The posterior mean of the spatial autocorrelation for the SDM differs but this may be because Bivand *et al.* [11] constrain ρ to be in the (0,1) interval. The SDM model seems to have a weaker residual spatial correlation, probably because the inclusion of the lagged covariates reduces the autocorrelation in the response.

Regarding INLA and MCMC estimates, although the marginals are close for the SLM model, they are a bit further away for all the other three models presented in Figure 6. The posterior modes are close but these differences may be due to the fact that two different link functions are used as the MCMC implementation defines a latent continuous variable y_i^* so that the response is 1 when y_i^* is non-negative and 0 otherwise. This makes the models fit with INLA and MCMC different in practice. The results obtained in the simulation study included in the Appendix also support that the differents are due to the different link functions.

	SEM	SLM	SDM	SDMlag	SDEM	SDEMlag	SLX	SLXlag
(Intercept)	-19.800	-8.142	-15.108		-12.775		-9.116	
flood_depth	-0.434	-0.187	-0.480	0.012	-0.497	0.100	-0.323	0.048
log_medinc	1.925	0.776	2.329	-0.891	2.092	-0.887	1.453	-0.582
small_size	-0.348	-0.403	-0.431	-1.113	-0.403	-0.583	-0.287	-0.703
large_size	-0.403	-0.475	-0.372	-1.646	-0.416	-1.913	-0.242	-1.011
low_status_customers	-0.335	-0.480	-0.103	-1.759	-0.089	-1.620	-0.059	-1.083
high_status_customers	0.135	0.105	0.049	0.113	0.067	0.261	0.035	0.083
owntype_sole_proprietor	0.774	0.790	0.910	1.393	0.844	0.853	0.613	0.808
owntype_national_chain	0.087	0.138	0.213	2.066	0.174	2.449	0.119	1.167
DIC	664.563	664.358	683.732		664.680		703.411	
M. Lik.	-386.500	-401.648	-435.459		-408.367		-410.769	

Table 2: Summary of point estimates of the fixed effects coefficients, Katrina business data.

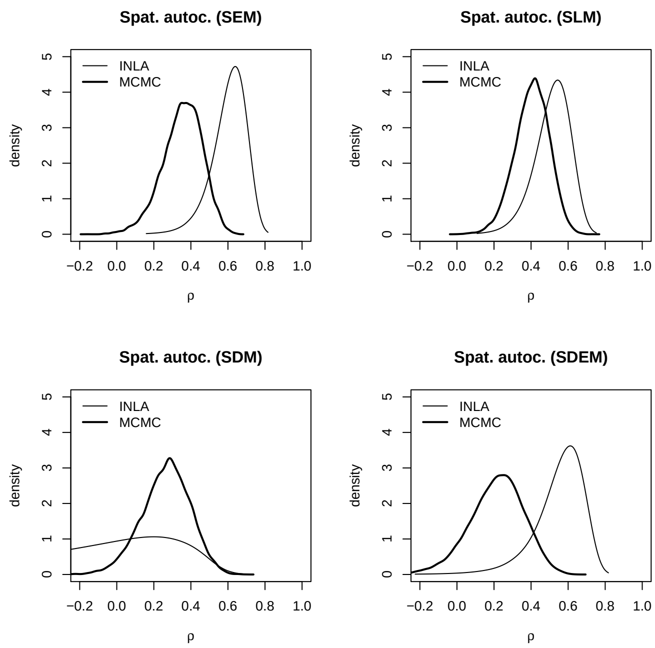

Figure 6. Posterior marginal of the spatial autocorrelation parameters for different models for INLA and MCMC, Katrina business data. Differences in the estimates can be explained by the different link functions used in the models fit with MCMC and INLA.

7.2.2. Exploring the number of neighbours

LeSage *et al.* [21] use a nearest neighbours method to obtain an adjacency matrix for the businesses in the dataset. Also, they have explored the optimal number of nearest

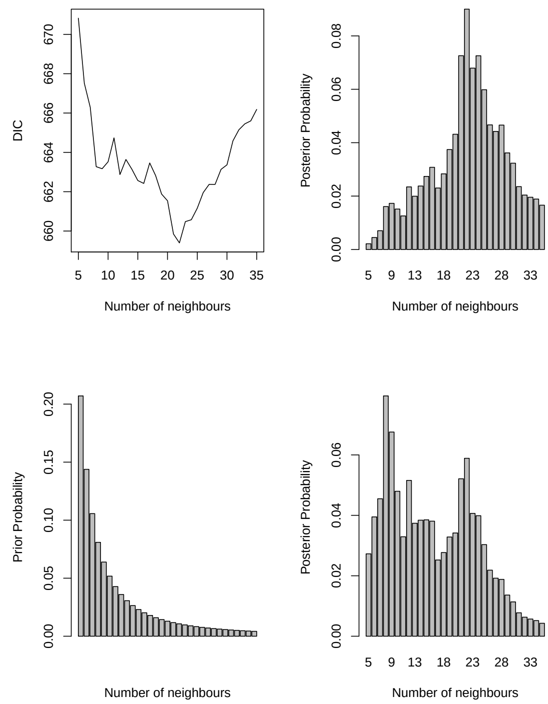

Figure 7. DIC (top-left) and posterior probabilities (top-right) for spatial probit models with different adjacency structures based on a nearest neighbours approach using a uniform prior, Katrina business data. Prior (bottom-left) and posterior (bottom-right) probabilities for spatial probit models with different adjacency structures based on a nearest neighbours approach using an informative prior, Katrina business data.

neighbours by fitting the model using different numbers of nearest neighbours and using the model with the lowest DIC [26] as the one with the optimal number of neighbours. For the 3-month horizon model, they compared a window of 8-14 neighbours, probably because of the computational burden of MCMC, which they used to fit their models.

The newly available `slm` model in **R-INLA** makes exploring the optimal number of neighbours faster and simpler. We have increased the number of neighbours considered, between 5 and 35, and fitted the SLM model using different adjacency matrices. These adjacency matrices have been created using the nearest neighbour algorithm with different values of the number of neighbours. Figure 7 shows how the optimal number of neighbours seems to be 22 according to both the DIC and the posterior probability (as explained in Section 6.1) criteria. However, we believe that this should be used as a guidance to set the number of neighbours as there may be other factors to take into account. In particular, a nearest neighbour approach may consider as neighbours businesses that are in different parts of the city, particularly if the number of nearest neighbours is allowed to be high.

When computing the posterior probability, the prior probability of each model is taken so that $\pi(\mathcal{M}_i) \propto 1$, and there is no prior preference on the number of neighbours. If we decide to favour adjancencies based on a small number of neighbours we may use a more informative prior. For example, we could take $\pi(k) \propto \frac{1}{k^2}$ so that neighbourhoods with a smaller number of neighbours are preferred. Figure 7 shows the prior and posterior probability for different values of the number of neighbours. Now it can be seen how our prior information produces different posterior probabilities, with an optimal number of neighbours of 8.

Finally, it is also possible to average over the ensemble of models using Bayesian model averaging. This will produce a fitted model that takes into account all the adjacency structures, weighted according to their marginal likelihoods. Table 3 shows the estimates of the fixed effects for the model with the highest posterior probability according to a uniform prior ($\pi(k) \propto 1$) and an informative prior ($\pi(k) \propto 1/k^2$), and the estimates obtained by averaging over all the fitted models. These models should take into account the uncertainty about the number of neighbours and can provide different estimates of the fixed effects. As it happens, the posterior means and standard deviations are slightly different, but we can observe higher differences in the case of an informative prior.

	Model with highest post. prob				BMA models			
	Uniform prior		Informative prior		Uniform prior		Informative prior	
	mean	sd	mean	sd	mean	sd	mean	sd
(Intercept)	-6.831	2.880	-8.645	3.086	-7.452	3.109	-8.544	3.299
flood_depth	-0.130	0.054	-0.206	0.056	-0.155	0.067	-0.201	0.070
log_medinc	0.643	0.281	0.829	0.301	0.708	0.304	0.819	0.323
small_size	-0.409	0.187	-0.407	0.186	-0.421	0.188	-0.410	0.187
large_size	-0.419	0.441	-0.545	0.449	-0.457	0.445	-0.509	0.449
low_status_customers	-0.436	0.207	-0.492	0.208	-0.450	0.212	-0.497	0.213
high_status_customers	0.099	0.171	0.124	0.171	0.108	0.172	0.107	0.172
owntype_sole_proprietor	0.805	0.262	0.757	0.262	0.788	0.262	0.772	0.262
owntype_national_chain	0.085	0.503	0.107	0.496	0.106	0.500	0.116	0.496

Table 3: Summary statistics of the covariate coefficients for the model with the highest probability (under two different priors) and the BMA model (under two different priors.)

7.2.3. Impacts

We have followed the method described in Section 5 to approximate the impacts for the models fitted to the Katrina dataset. In this case, we do not have any model fitted using ML with which to compare. The implementation of the spatial probit in the Spatial Econometrics Toolbox is for a different link, so our results cannot directly be compared to MCMC as reported in LeSage *et al.* [21]. Direct impacts are shown in Table 4, whilst total and indirect impacts are available as supplementary materials.

	INLASEM	INLASLM	INLASDM	INLASDEM	INLASLX
flood_depth	-0.086	-0.038	-0.100	-0.099	-0.090
log_medinc	0.381	0.159	0.504	0.409	0.386
small_size	-0.070	-0.084	-0.097	-0.079	-0.079
large_size	-0.078	-0.100	-0.085	-0.080	-0.068
low_status_customers	-0.067	-0.100	-0.034	-0.018	-0.016
high_status_customers	0.027	0.022	0.009	0.014	0.010
owntype_sole_proprietor	0.154	0.165	0.201	0.165	0.169
owntype_national_chain	0.016	0.025	0.060	0.033	0.033

Table 4: Direct impacts, Katrina data set.

Inference on the impacts relies on sampling from the approximation to the joint posterior distribution which is then used to compute the impacts and estimate their posterior distributions, as we have already seen in the Boston housing data example. Figure 8 shows the estimates of the average total impacts for flood depth for four models. In a thicker line we have included the posterior marginal of the impacts computed using the output from MCMC using the **spatialprobit** package. As previously stated, MCMC results can be roughly compared to our SLM model estimates but keeping in mind that different link functions have been used and that differences may appear. In this case, the quality of our approximations differ with the models. We have found similar accuracy for all the other variables, which means that our approximation appears to be acceptable.

Note that the impacts estimated with INLA and MCMC are close regardless of the estimates of the spatial autocorrelation parameters shown in Figure 6. We believe that this is because the impacts themselves build on the fitted model parameter values rather than the posterior distributions, which only enter into simulations to get to the distributions of the impacts.

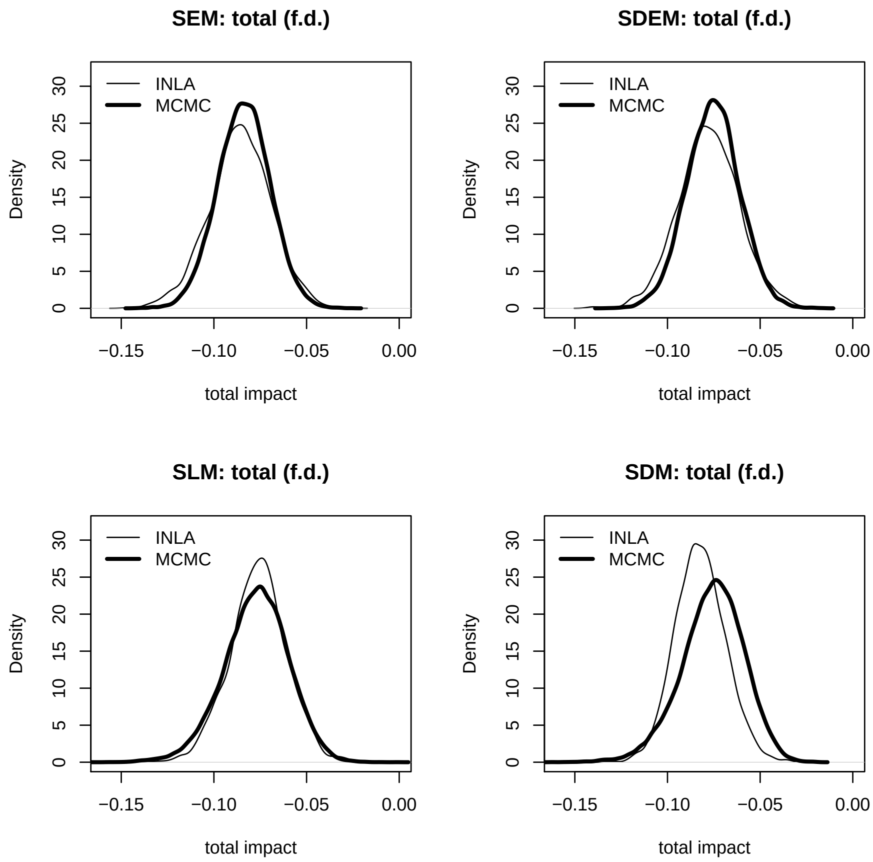

Figure 8. Average total impacts for flood depth for the Katrina dataset.

8. Discussion and final remarks

In this paper we have described how the analysis of spatial econometrics data requires the use of very specific models and how the integrated nested Laplace approximation offers an alternative to model fitting. Instead of resorting to MCMC methods, INLA aims at providing approximate inference on the marginal distribution of the model parameters. This methodology is implemented in the R package **R-INLA**, which includes a particular latent effect called `slm` which can be used to fit many spatial econometrics models.

We have also shown how impacts can be approximated when models are fitted with **R-INLA**. As this requires multivariate posterior inference, the estimation of the impacts is achieved by sampling from the approximation to the joint posterior of the required coefficients and spatial autocorrelation parameter and computing the impacts using these samples. It is worth noting that these samples are independent, so less samples are required for inference than with typical MCMC algorithms.

It should be noted that are several the advantages of using INLA and **R-INLA**. Not only model specification and fitting is very easy using R but also the computational speed allows us to explore a large number of models. When the model is not available within the range of latent models available in **R-INLA** it is often possible to fit conditional models, on one or two model parameters, and then obtained the desired model by averaging over these models. Furthermore, other important topics in Bayesian inference, such as prediction of missing responses, model selection and variable selection can be tackled with INLA. Several authors [see, for example, 31,32] have assessed the accuracy of INLA (as compared to MCMC) for a wide range of models, so we believe that this will also be the case for the spatial econometrics models presented herein.

In the future, we expect to explore how to increase the number of Spatial Econometrics models available in **R-INLA** and how to extend them. In particular, we find that there is interesting work to do on models with more than one spatial term, spatio-temporal models, the analysis of panel data and how to account for measurement error in covariates, for example.

Author Contributions: All authors contributed to the conceptualization, writing and editing of this article. They have read and agreed to the published version of the manuscript.

Funding: V. Gómez-Rubio has been supported by grants SBPLY/17/180501/000491 (Consejería de Educación, Cultura y Deportes, JCCM, Spain, and FEDER) and PID2019-106341GB-I00 (Ministerio de Ciencia e Innovación, Spain).

Data Availability Statement: Data are available from different R packages (available from CRAN https://cran.r-project.org) and the scripts to reproduce the results in the paper are at https://github.com/becarioprecario/slm.

Conflicts of Interest: The authors declare no conflict of interest.

1. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **2009**, *71*, 319–392. doi:{10.1111/j.1467-9868.2008.00700.x}.
2. Hepple, L. Bayesian analysis of the linear model with spatial dependence. In *Exploratory and explanatory statistical analysis of spatial data*; Bartels, C.P.A.; Ketellapper, R.H., Eds.; Martinus Nijhoff: Boston, 1979; pp. 179–199.
3. Anselin, L. A note on small sample properties of estimators in a first-order spatial autoregressive model. *Environment and Planning A* **1982**, *14*, 1023–1030.
4. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer: Dordrecht, 1988.
5. Hepple, L.W. Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds. *Environment and Planning A* **1995**, *27*, 447–469.
6. Hepple, L.W. Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms. *Environment and Planning A* **1995**, *27*, 615–644.
7. Hepple, L. Bayesian model choice in spatial econometrics. In *Spatial and spatiotemporal econometrics*; LeSage, J.P.; Pace, R.K., Eds.; Elsevier: Amsterdam, 2004; Vol. 18, *Advances in Econometrics*, pp. 101–126.
8. LeSage, J.P. Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review* **1997**, *20*, 113–129.
9. LeSage, J.P. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis* **2000**, *32*, 19–35.
10. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; Chapman and Hall/CRC, 2009.
11. Bivand, R.S.; Gómez-Rubio, V.; Rue, H. Approximate Bayesian inference for spatial econometrics models. *Spatial Statistics* **2014**, *9*, 146 – 165. Revealing Intricacies in Spatial and Spatio-Temporal Data: Papers from the Spatial Statistics 2013 Conference, doi: http://dx.doi.org/10.1016/j.spasta.2014.01.002.
12. Bivand, R.S.; Gómez-Rubio, V.; Rue, H. Spatial Data Analysis with R-INLA with Some Extensions. *Journal of Statistical Software* **2015**, *63 (20)*.
13. Hoeting, J.; David Madigan and, A.R.; Volinsky, C. Bayesian Model Averaging: A Tutorial. *Statistical Science* **1999**, *14*, 382–401.
14. Gómez-Rubio, V.; Palmí-Perales, F. Multivariate posterior inference for spatial models with the integrated nested Laplace approximation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **2019**, *68*, 199–215, [https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12292]. doi:10.1111/rssc.12292.
15. Gómez-Rubio, V.; Bivand, R.S.; Rue, H. Bayesian Model Averaging with the Integrated Nested Laplace Approximation. *Econometrics* **2020**, *8*. doi:10.3390/econometrics8020023.
16. Anselin, L. Thirty years of spatial econometrics. *Papers in Regional Science* **2010**, *89*, 3–25.
17. LeSage, J.P.; Pace, R.K. Spatial Econometric Models. In *Handbook of Applied Spatial Analysis*; Fischer, M.; Getis, A., Eds.; Springer-Verlag: Heidelberg, 2010; pp. 355–376.
18. Haining, R. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press, 2003.

19. Pace, R.K.; LeSage, J.P.; Zhu, S. Spatial dependence in regressors and its effect on performance of likelihood-based and instrumental variable estimators. Advances in Econometrics; Terrell, D.; Millimet, D., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2012; Vol. 30, pp. 257–295.
20. Rue, H.; Held, L. *Gaussian Markov Random Fields. Theory and Applications*; Chapman & Hall/CRC, 2005.
21. LeSage, J.P.; Pace, K.R.; Lam, N.; Campanella, R.; Liu, X. New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **2011**, *174*, 1007–1027. doi:10.1111/j.1467-985X.2011.00712.x.
22. Billé, A.G. Computational Issues in the Estimation of the Spatial Probit Model: A Comparison of Various Estimators. *The Review of Regional Studies* **2013**, *43*, 131–154.
23. Calabrese, R.; Elkink, J.A. ESTIMATORS OF BINARY SPATIAL AUTOREGRESSIVE MODELS: A MONTE CARLO STUDY. *Journal of Regional Science* **2014**, *To appear*.
24. Besag, J.; York, J.; Mollie, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **1991**, *43*, 1–59. doi:10.1007/BF00116466.
25. Gómez-Rubio, V. *Bayesian inference with INLA*; CRC Press: Boca Raton, FL, 2020.
26. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van der Linde, A. Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society, Series B* **2002**, *64*, 583–616.
27. Roos, M.; Held, L. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* **2011**, *6*, 259–278. doi:{10.1214/11-BA609}.
28. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **2006**, *3*, 515–533.
29. Harrison, D.; Rubinfeld, D.L. Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Management* **1978**, *5*, 81–102.
30. Pace, R.K.; Gilley, O.W. Using the Spatial Configuration of the Data to Improve Estimation. *Journal of the Real Estate Finance and Economics* **1997**, *14*, 333–340.
31. De Smedt, T.; Simons, K.; Van Nieuwenhuyse, A.; Molenberghs, G. Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. *Archives of Public Health* **2015**, *O2*. doi:https://doi.org/10.1186/2049-3258-73-S1-O2.
32. Grilli, L.; Metelli, S.; Rampichini, C. Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models. *Journal of Statistical Computation and Simulation* **2015**, *13*, 2718–2726. doi:10.1080/00949655.2014.935377.
33. LeSage, J.P.; Pace, R.K. Spatial econometric Monte Carlo studies: raising the bar. *Empirical Economics* **2018**, pp. 17–34.

Appendix A R interface

The implementation of the `slm` latent model in **R-INLA** requires careful attention to the different parameters included in the model. It can be defined using the `f()` function in **R-INLA** as

```
f(idx, model="slm", args.slm=list(rho.min , rho.max, W, X, Q.beta), hyper)
```

`idx` is an index to identify the region, and it can take any values, `model="slm"` indicates that we are using a `slm` latent effect, `args.slm` sets some of the data required to fit the model (including the prior on β) and `hyper` is used to set the prior distributions for τ and ρ.

`rho.min` and `rho.max` indicate the range of the spatial autocorrelation parameter ρ. This range will depend on the spatial weights matrix defined in `W`. See for example Haining [18] for details.

`W` is an adjacency matrix that defines the spatial structure of the data. In spatial econometrics `W` is often taken to be row-standardised. **R-INLA** can handle sparse matrices, as defined in the **Matrix** package, to make model fitting faster.

`X` is the matrix of column-wise covariates. If an intercept is required in the model, it must be included as a column of 1's (possibly, in the first column). `Q.beta` is a precision matrix for the vector of coefficients β. In principle, it can take any form but we advise taking diagonal matrix with very small values in the diagonal.

In order to set $\beta = 0$ we have considered a model with a matrix with zero columns in R. This will mimic a model with no covariates at all.

If the covariates have very different scales it may important to re-scale them. Otherwise, **R-INLA** may find some computational problems that may prevent it from fitting the model. This is particularly important for large datasets.

`hyper` can be used to assign prior distributions to τ and ρ in the same way as the `hyper` parameter in the `f()` functions and that is described in the **R-INLA** documentation.

The posterior marginal of ρ is reported (and constrained) on the interval (0,1) and not in the range defined by `rho.min` and `rho.max`. Hence, in order to make an appropriate interpretation of the results it must be linearly transformed to fit in the (`rho.min`, `rho.max`) interval using, for example, `inla.tmarginal()`. Note that if an initial value is assigned to ρ this must be re-scaled to be in the (0,1) interval as well.

We will not include here any example with R code on the use of this new latent class. The R files used to calculate the examples in Section 7 are distributed as supplementary materials to this paper.

Appendix B Expression of `slm` as a Gaussian Markov Random Field

In this Appendix we will show how the newly defined `slm` latent effect can be expressed as a Gaussian Markov Random Field. We will denote the vector of random effects as

$$\mathbf{x} = (I_n - \rho W)^{-1}(X\beta + \varepsilon)$$

Here β has a Gaussian prior with zero mean and precision matrix Q, and ε a Gaussian distribution with zero mean and precision matrix τI_n, with τ a precision parameter. Q is fixed when the latent effects are defined, so we will treat it as constant. Although not explicitly written down, we are also conditioning on hyperparameters τ and ρ in all the distributions that appear below.

Internally, **R-INLA** works with the joint distribution of $\mathbf{x}$ and β, denoted by $[\mathbf{x}, \beta]$. We will show here that this can be expressed as a GMRF with a sparse precision matrix, so that it conforms with the INLA framework.

First of all, we will work out the conditional distribution of $\mathbf{x}$ on β, denoted by $[\mathbf{x}|\beta]$. We will use this later because $[\mathbf{x}, \beta] = [\mathbf{x}|\beta][\beta]$.

We are assuming that the joint distribution is Gaussian and, hence, the conditional distribution $[\mathbf{x}|\beta]$ is also Gaussian, with

$$M = E[\mathbf{x}|\beta] = (I_n - \rho W)^{-1}X\beta$$

and

$$\begin{aligned} Var[\mathbf{x}|\beta] = Var[(I_n - \rho W)^{-1}X\beta + (I_n - \rho W)^{-1}\varepsilon|\beta] = \\ (I_n - \rho W)^{-1}Var[\varepsilon|\beta]((I_n - \rho W)^{-1})' = \\ (I_n - \rho W)^{-1}\frac{1}{\tau}I_n((I_n - \rho W)^{-1})' = \\ \frac{1}{\tau}(I_n - \rho W)^{-1}(I_n - \rho W')^{-1} \end{aligned} \tag{A32}$$

The conditional precision can be expressed as

$$T = Prec[\mathbf{x}|\beta] = \tau(I_n - \rho W')(I_n - \rho W)$$

which is symmetric and highly sparse.

Hence, we can derive the joint distribution of $\mathbf{x}$ and β as

$$\begin{aligned} [\mathbf{x}, \beta] = [\mathbf{x}|\beta][\beta] \propto \exp\{-\frac{1}{2}(\mathbf{x} - M)'T(\mathbf{x} - M)\}\exp\{-\frac{1}{2}\beta'Q\beta\} = \\ \exp\{-\frac{1}{2}\Big(\mathbf{x}'T\mathbf{x} - \mathbf{x}'TM - M'T\mathbf{x} + M'TM + \beta'Q\beta\Big)\} = \\ \exp\{-\frac{1}{2}(\mathbf{x}, \beta)'P(\mathbf{x}, \beta)\} \end{aligned} \tag{A33}$$

Here P is the precision matrix of $[\mathbf{x}, \beta]$, which is given by

$$\begin{aligned} P = \begin{bmatrix} T & -T(I_n - \rho W)^{-1}X \\ -X'(I_n - \rho W')^{-1}T & Q + \tau X'X \end{bmatrix} = \\ \begin{bmatrix} \tau(I_n - \rho W')(I_n - \rho W) & -\tau(I_n - \rho W')X \\ -\tau X'(I_n - \rho W) & Q + \tau X'X \end{bmatrix} \end{aligned} \tag{A34}$$

Note that to obtain the previous result we have used that

$$\mathbf{x}'TM = \mathbf{x}'T(I_n - \rho W)^{-1}X\beta = \\ \mathbf{x}'(\tau(I_n - \rho W')(I_n - \rho W))(I_n - \rho W)^{-1}X\beta = \\ \tau\mathbf{x}'(I_n - \rho W')X\beta, \tag{A35}$$

$$M'T\mathbf{x} = (\mathbf{x}'TM)' = \tau\beta'X'(I_n - \rho W)\mathbf{x},$$

and

$$M'TM = \beta'X'(I_n - \rho W')^{-1}T(I_n - \rho W)^{-1}X\beta = \\ \beta'X'(I_n - \rho W')^{-1}(\tau(I_n - \rho W')(I_n - \rho W))(I_n - \rho W)^{-1}X\beta = \\ \tau\beta'X'X\beta \tag{A36}$$

Furthermore, from the final expression in equation (A33) it is easy to see that the expectation of $[\mathbf{x}, \beta]$ is

$$E[\mathbf{x}, \beta] = 0$$

Hence, the expression of $[\mathbf{x}, \beta]$ as a Gaussian Markov Random Field has zero mean and precision matrix P. Note how P is a block-matrix which involves very sparse matrices, which allows for the use of the efficient algorithms described in Rue and Held [20] for fast computation on GMRF.

Although Q can take any form, assuming that the coefficients are independent a priori will lead to a diagonal matrix, which is also sparse. Finally, it may be worth rescaling the covariates to avoid numerical problems. Including lagged covariates may lead to further numerical instability as they may be highly correlated with the original covariates.

Supplementary materials

We have prepared a number of R files to be distributed with this paper. These files show how to obtain the results that we have presented here. Note that all these examples require the use of several R packages. Data are available from different R packages and the scripts to reproduce the results in the paper are at https://github.com/becarioprecario/slm. We provide here a list of these files and a short description:

- `boston-slm.R`
  Analysis of the Boston housing data set using the main spatial econometrics models.
- `boston-slm-impacts.R`
  Computation of the impacts for the Boston housing data example.
- `boston-slm-full.R`
  Analysis of the Boston housing data set using the main spatial econometrics models and the full adjacency matrix to perform prediction on the missing values.
- `katrina-slm.R`
  Analysis of the Katrina business data using the main spatial econometrics models with a spatial probit.
- `katrina-slm-neigh.R`
  Selection of the number of optimal nearest neighbours for the adjacency matrix using the Katrina business data.
- `katrina-slm-impacts.R`
  Computation of the impacts for the Katrina business data example.
- `simulation_study.R`
  Code used in the simulation study provided in the Appendix.

Boston housing data: supplementary materials

	SEM	SLM	SDM	SDEM
mean	0.745	0.497	0.613	0.674
sd	0.034	0.031	0.042	0.043
0.025 quantile	0.675	0.432	0.531	0.585
0.975 quantile	0.807	0.555	0.696	0.754

Table 5: Summary of the spatial autocorrelation parameters, Boston housing data.

	MLSLM	INLASLM	MLSDM	INLASDM	INLASDEM	INLASLX
CRIM	-0.008	-0.008	-0.007	-0.007	-0.007	-0.007
ZN	0.000	0.000	0.000	0.000	0.000	0.000
INDUS	0.002	0.002	0.000	0.000	0.001	-0.002
CHAS1	-0.003	-0.002	-0.049	-0.050	-0.047	-0.065
I(NOX^2)	-0.246	-0.249	-0.061	-0.060	-0.050	0.081
I(RM^2)	0.008	0.008	0.010	0.010	0.010	0.009
AGE	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001
log(DIS)	-0.149	-0.149	-0.056	-0.050	-0.082	-0.038
log(RAD)	0.066	0.066	0.057	0.057	0.059	0.043
TAX	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
PTRATIO	-0.013	-0.013	-0.015	-0.015	-0.016	-0.012
B	0.000	0.000	0.001	0.001	0.000	0.001
log(LSTAT)	-0.231	-0.231	-0.229	-0.230	-0.233	-0.234

Table 6: Direct impacts, Boston housing data.

	MLSLM	INLASLM	MLSDM	INLASDM	INLASDEM	INLASLX
CRIM	-0.014	-0.014	-0.024	-0.024	-0.013	-0.022
ZN	0.001	0.001	0.001	0.001	-0.000	0.001
INDUS	0.004	0.004	0.001	0.000	0.001	0.000
CHAS1	-0.005	-0.005	0.127	0.127	0.086	0.129
I(NOX^2)	-0.461	-0.467	-1.033	-1.029	-0.634	-0.975
I(RM^2)	0.015	0.015	0.004	0.003	0.009	0.003
AGE	-0.000	-0.000	0.001	0.001	-0.000	0.001
log(DIS)	-0.279	-0.282	-0.273	-0.274	-0.158	-0.259
log(RAD)	0.123	0.124	0.125	0.126	0.073	0.121
TAX	-0.001	-0.001	-0.000	-0.000	-0.000	-0.000
PTRATIO	-0.025	-0.025	-0.039	-0.039	-0.031	-0.038
B	0.000	0.000	0.000	0.000	0.000	-0.000
log(LSTAT)	-0.434	-0.434	-0.424	-0.425	-0.345	-0.414

Table 7: Total impacts, Boston housing data.

	MLSLM	INLASLM	MLSDM	INLASDM	INLASDEM	INLASLX
CRIM	-0.007	-0.007	-0.016	-0.016	-0.006	-0.015
ZN	0.000	0.000	0.000	0.000	-0.001	0.000
INDUS	0.002	0.002	0.001	0.000	0.001	0.002
CHAS1	-0.002	-0.002	0.176	0.177	0.133	0.194
I(NOX^2)	-0.216	-0.218	-0.973	-0.969	-0.585	-1.056
I(RM^2)	0.007	0.007	-0.006	-0.007	-0.002	-0.007
AGE	-0.000	-0.000	0.003	0.003	0.001	0.002
log(DIS)	-0.130	-0.132	-0.218	-0.224	-0.076	-0.221
log(RAD)	0.058	0.058	0.068	0.069	0.014	0.078
TAX	-0.000	-0.000	0.000	0.000	0.000	0.000
PTRATIO	-0.012	-0.012	-0.024	-0.024	-0.016	-0.027
B	0.000	0.000	-0.000	-0.000	-0.000	-0.001
log(LSTAT)	-0.203	-0.203	-0.195	-0.196	-0.112	-0.180

Table 8: Indirect impacts, Boston housing data.

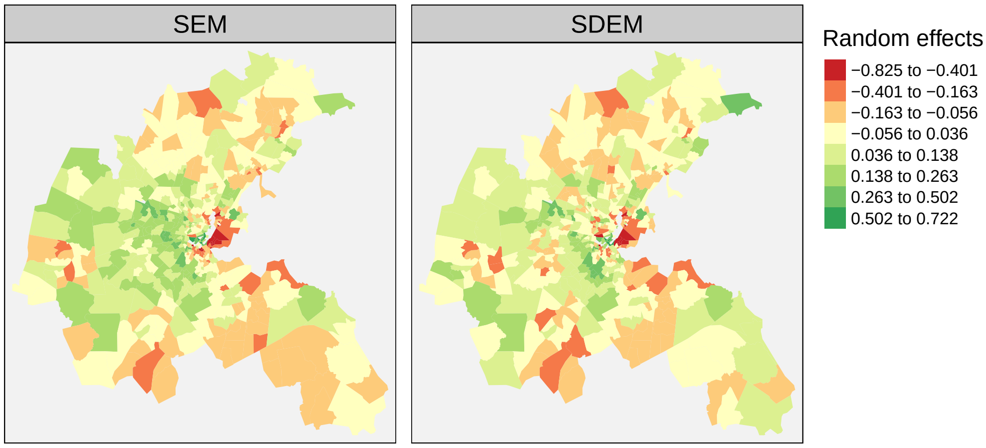

Figure A9. Posterior means of the `slm` latent effects for the SEM and SDEM models, Boston housing data with full adjacency matrix.

Katrina business data: supplementary materials

	SEM	SLM	SDM	SDEM
mean	0.605	0.513	-0.113	0.550
sd	0.095	0.098	0.434	0.137
0.025 quantile	0.381	0.290	-1.115	0.209
0.975 quantile	0.753	0.676	0.505	0.744

Table 9: Summary of point estimates of the spatial autocorrelation parameters, Katrina business data.

	INLASEM	INLASLM	INLASDM	INLASDEM	INLASLX
flood_depth	-0.086	-0.076	-0.083	-0.078	-0.076
log_medinc	0.381	0.316	0.250	0.239	0.242
small_size	-0.070	-0.173	-0.287	-0.192	-0.271
large_size	-0.078	-0.208	-0.371	-0.446	-0.347
low_status_customers	-0.067	-0.201	-0.339	-0.333	-0.321
high_status_customers	0.027	0.046	0.028	0.059	0.029
owntype_sole_proprietor	0.154	0.338	0.420	0.336	0.392
owntype_national_chain	0.016	0.050	0.406	0.506	0.353

Table 10: Total impacts, Katrina data set.

	INLASEM	INLASLM	INLASDM	INLASDEM	INLASLX
flood_depth	0.000	-0.037	0.018	0.021	0.014
log_medinc	0.000	0.157	-0.254	-0.169	-0.144
small_size	0.000	-0.088	-0.190	-0.113	-0.192
large_size	0.000	-0.108	-0.286	-0.366	-0.279
low_status_customers	0.000	-0.101	-0.305	-0.316	-0.304
high_status_customers	0.000	0.024	0.019	0.045	0.018
owntype_sole_proprietor	0.000	0.173	0.219	0.170	0.223
owntype_national_chain	0.000	0.026	0.346	0.473	0.320

Table 11: Indirect impacts, Katrina data set.

Simulation study

In order to assess that INLA provides good estimates of the model parameters a simulation study has been conducted. In particular, given that the aim is to assess the performance of the newly implemented `slm` latent effect when estimating the model parameters, data will be simulated using a spatial lag model (SLM) as this is the more general model. The SEM, SDM and SDEM can be regarded as particular cases or extensions of the SLM model.

When simulating the data we will be relying on the adjacency structure of the Boston housing data. We will consider the full adjacency structure with the original 506 areas. In addition, the model will include a covariate, which has been simulated using a uniform distribution between -3 and 3 to provide an ample range of values. The value of the intercept is -1 and the value of the coefficient of the covariate is 1. Furthermore, the value of the precision of the latent effects is 1 and the value of the spatial autocorrelation parameter is 0.5, so that data are simulated with a positive spatial autocorrelation.

A total of 100 simulated datasets have been generated under this model using a Gaussian reponse. Similarly, in order to simulate the data to fit a spatial probit model, the linear predictors obtained in the previous simulations have been used to compute the probabilities (by taking the inverse of the probit function) and then these probabilities have been used to simulate the outcome (which is a binary variable) using a Bernoulli distribution.

In order to assess how accurate the estimates are, three different criteria have been used. The mean absolute error (MAE) and mean relative error (MRE) have been computed (using the posterior means of the model parameters) provide point-based error criteria, while the percentage of coverage using 95% credible intervals (i.e., the percentage of times the credible interval contains the actual value of the parameter) will provide a quality criterion that considers the uncertainty of the estimates. All these results are available in Table 12.

Parameter	Model					
	Gaussian			Bernoulli		
	MAE	MRE	Coverage	MAE	RAE	Coverage
Intercept	0.14	0.14	0.89	0.11	0.11	0.98
Coefficient	0.03	0.03	0.89	0.06	0.06	0.97
Precision	0.07	0.07	0.84	–	–	–
Spatial autocorrelation	0.06	0.12	0.90	0.06	0.11	0.97

Table 12: Summary of the different criteria used to assess the quality of the estimates using the simulated data.

In general, the percentage of coverage is quite good for both models, and close to the nominal 95%. MAE and RAE are also good, with the highest relative error close to 14% (for the coefficient of the covariate).

LeSage and Pace [33] provide a thorough discussion about the design of Monte Carlo studies for spatial econometrics models. We have provided a simpler Monte Carlo simulation study here as the aim here is to show for a simple model that INLA is able to recover the values of the main model parameters. A more thorough simulation study (that may consider edge cases such as values of the spatial autocorrelation parameter close to 1) is out of the scope of this paper.