Approximate Bayesian Inference for Spatial Econometrics Models

Roger S. Bivand, Virgilio Gómez-Rubio and Håvard Rue

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 very computationally intensive simulation methods, such as Markov Chain Monte Carlo. However, when only marginal inference is needed, INLA provides a fast and accurate estimate of the posterior marginals of the parameters in the model.

Futhermore, we have compared the results provided by these models to those obtained with a more general class of Generalized 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 and then varying the values of the fixed parameters. The posterior marginal of the parameters of interest is 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 parametes. This approach can be used to fit even more general models.

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

1 Introduction

Economic data often shows spatial patterns. For example, flat prices are similar in adjacent neighbourhoods, GDP varies smoothly accross countries, etc. Spatial Econometrics models aim at including this spatial dependence so that the observed value in an area depends on the observed values at the neighbours.

Traditionally, for a continous observed variable (e.g., flat prices) this dependance has been done explicitedly, i.e., the price in an area is centred at a weighted average of the observed values at the neighbours plus perhaps the effect of other covariates. Several autoregressive models have been built on this idea, as described in Section 2.1. However, when the response is non-Gaussian autoregressive models are difficult to handle. This is discussed in Section 2.2.

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. Hence, spatial correlation between observations is obtained as a result of the correlation of the latent random effect.

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 intesive methods such as Markov Chain Monte Carlo to obtain the joint posterior distribution of the parameters. Once this has been obtained, it is very easy to compute summary statistics of the model parameters, credible intervals and other quantities of interest.

However, when only marginal inference is needed other methods are available. Rue et al. (2009) describe the Integrated Nested Laplace Approximation (INLA) to obtain an approximation to the posterior marginals of the parameters of interest. They also provide software to fit a wide range of models which in most cases reduces computation time to seconds and allows for the use of larger data sets.

Our aim is to apply this new methodology to Spatial Econometrics models. We have considered four models that are widely used nowadays. For Gaussian data, we will work with the Spatial Autoregressive, Spatial Lag and Spatial Durbin models. For binary data, we will consider the Spatial Probit, which relies on the Spatial Autoregressive model. LeSage and Pace (2009) provide a full description of the most important models in Spatial Econometrics and software for Bayesian inference in their Spatial Econometrics Tooldbox (`http://www.spatial-econometrics.com/`).

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 the Integrated Nested Laplace Aproximation 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. Finally, Section 7 includes a general discussion of the paper.

2 Spatial Econometrics Models

{sec:SEM}

Spatial models have been used in spatial econometrics for a long time (see Anselin, 2010, for a review). In general, the interest in Spatial Econometrics is on modeling spatial interaction in an autoregressive way, so that the observation at a given area, y_i, depends on a weighted sum on the the values of the variable at the 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

{subsec:SARs}

A popular model is the Simultaneous Autoregressive model, which can be written down as follows:

$$y = X\beta + u; u = \rho W u + e; e \sim N(0, \sigma^2)$$

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 covariates coefficients and W is a row-standardised adjacency matrix. ρ is a parameter that measures spatial autocorrelation and, given that W is row-standardised, it is bound to be between -1 and 1 (Haining, 2003). Furthermore, the error terms are modelled to be a weighted sum of the random errors at the neighbours plus some random noise. This model can be re-written as

$$y = X\beta + e'; e' \sim N(0, \sigma^2 (I - \rho W)^{-1} (I - \rho W')^{-1})$$

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

$$y = \rho W y + X\beta + e; e \sim N(0, \sigma^2)$$

This model can be rewritten as:

$$y = (I - \rho W)^{-1} X\beta + e';\ e' \sim N(0, \sigma^2 (I - \rho W)^{-1} (I - \rho W')^{-1})$$

Finally, a third model that is widely used in Spatial Econometrics is the Spatial Durbin model:

$$y = \rho W y + X\beta + WX\gamma + e = [X, WX][\beta, \gamma] + e;\ ; e \sim N(0, \sigma^2)$$

γ is a coefficient for the lagged covariates WX, so that the response now depends not only on the response at the neighbours and the covariates, but on the (weighted) covariates at the neighbours.

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

$$y = \rho W y + XWX\beta' + e;\ XWX = [X, WX];\ \beta' = [\beta, \gamma]$$

and

$$y = (I - \rho W)^{-1} XWX\beta' + e';\ e' \sim N(0, \sigma^2 (I - \rho W)^{-1}(I - \rho W')^{-1})$$

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.

LeSage and Pace (2009) provide Matlab code to fit all these models and obtain estimates and summary statistics of the model parameters. However, using MCMC can sometimes be very computationally intensive and simulations may take a while.

It should be noted that, 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 and γ. This is a general approach which could be used for other models that cannot be fitted with R-INLA but that become "R-INLA fitable" once some parameters (usually, one or two) have been fixed.

2.2 Non-Gaussian models

{subsec:probit}

When the response variable is binary (i.e., the outcome is zero or one), the previous models are not adecuate. LeSage et al. (2011) 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:

$$y_i = \begin{cases} 1 & if\ y_i^* \geq 0 \\ 0 & if\ y_i^* < 0 \end{cases} \quad (1)$$ {eq:probit}

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

$$y^* = \rho W y^* + X\beta + \varepsilon;\ \varepsilon \sim N(0, \sigma^2)$$

We will use the following expression to represent this model

$$y^* = (I - \rho W)^{-1} X\beta + e';\ e' \sim N(0, \sigma^2 (I - \rho W)^{-1} (I - \rho W')^{-1})$$

as it is easier to link with the Generalized Linear Models described below. Also, it is clear form equation (1) in this model that the relationship between y_i and y_i^* is non-linear.

2.3 Generalized Linear Models with random effects

A different way of modelling the outcome and accounting for covariates and spatial autocorrelation is by means of Generalized Linear Models **ref**. Outcome y_i is supposed to come from a distribution of the Exponential family with mean parameter μ_i. The relationship between μ_i and the linear predictor on the covariates is establisehd through a link function $g(\cdot)$:

$$g(\mu_i) = \eta_i = X_i \beta$$

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

$$\eta_i = X_i \beta + u_i$$

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

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

where I_n is the identity matrix of dimension $n \times n$.

Another widely used specification is the Conditionally autoregressive specification:

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

Model fitting and interpretation of this kind of models is somewhat easier than 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

$$(I - \rho W)X\beta$$

which clearly shows that the effect of a covariate depends on its coefficient and the spatial correlation ρ.

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

$$\eta_i = X\beta + WX\gamma + u_i$$

WRITE ABOUT THE IMPACTS???

It should be noted that the models described in Sections 2.1 and 2.2 can be expressed as Genelized Linear Models. The Gaussian models described in Section 2.1 are Gaussian GLMs with random effects after conditioning on ρ. The variance-covariante 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.

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

Bayesian inference for these models requires the use of Markov Chain Monte Carlo techniques. LeSage and Pace (2009) provide Matlab code to fit these and many other models in their Spatial Econometrics Toolbox. However, running this code for large data sets can be very time consuming and our aim is to find alternative ways of providing Bayesian inference. For this reason we will rely on th Integrated Nested Laplace Approximation, which is discussed in the next Section.

2.4 Impacts

LeSage and Pace (2009) and LeSage et al. (2011) discuss how changes of a covariate at location i will affect the output at location j. They define the *impacts* and *direct* and *indirect* effects to measure these effects.

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:

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

The matrix of impacts for covariate v can be derived as

$$\frac{\partial \eta}{\partial x'_v} = (I_n - \rho W)^{-1}\beta_v;\ v = 1, \dots, p$$

The *direct effects* associated to covariate v are 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 *indirect effects* are defined as the average 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 Binomial models is slightly different. Now the impacts are based on computing the partial derivatives of

$$\frac{\partial Pr(y_i = 1)}{\partial x'_{v,j}}$$

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 is

$$\frac{\partial Pr(y = 1)}{\partial x'_v} = D(f(\eta))(I_n - \rho W)^{-1}\beta_v;\ v = 1, \dots, p$$

Here $f(\eta)$ is a vector of the standard Normal distribution evaluated at the values of the linear predictors $\eta_i, i = 1, \dots, n$. $D(\cdot)$ simply represents a diagonal matrix made from its argument. Direct and indirect effects are defined from the matrix of partial derivatives as in the Gaussian case. See LeSage et al. (2011) 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.

$- - -- >$SOMETHING ABOUT THE IMPACTS

The posterior distribution of the impacts can be esaily 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. Unfortunately, this distribution is hard to compute when INLA is used for the model fitting. As a crude alternative, we compute an Empirical Bayes estimate of the impacts by plugging in the posterior means of the quantities required to compute them. Credible intervals cannot be provided though.

3 Approximate Inference using the Integrated Laplace Approximation

{sec:INLA}

Rue et al. (2009) and Lindgren et al. (2011) have develop an approximate method for Bayesian inference based on focusing on the marginals of the models parameters. 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 $\mathbf{y} = (y_1, \dots, y_n)$ the distribution of y_i is supposed to be one of the exponential family with mean μ_i. The relation between μ_i and a linear predictor of some latent effects is done via a link function. Note that 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 some 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 precission 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

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

where x_i is the latent linear predictor η_i and $\mathcal{I}$ represents the indices of the observations. Note that some of the value in $\mathbf{y}$ amy be missing and this is why the product is over a set of indices $\mathcal{I}$ and not form 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:

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

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

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

and

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

Rue et al. (2009) 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:

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

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

Rue et al. (2009) 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` (R Development Core Team, 2011) 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 Gomez-Rubio et al. (2012).

4 Model fitting with INLA

{sec:INLAext}

At the moment, the 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, in the Spatial Lag and Spatial Durbin models the linear term on the covariates is multiplied by $(I - \rho W)$, which is clearly not a standard linear term.

However, after conditioning on a value of ρ these models belong to the class of models that R-INLA can fit. However, 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

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

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 $\pi(y)$. It should be noted that the marginal distribution of ρ is

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

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

It shold be noted that, as W is row-standardised, ρ is in the range $(-1, 1)$. Hence, we could 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 approxiamte the marginal distribution of ρ.

For each value of ρ_j we can compute $\pi(y|\rho_j)\pi(\rho_j)$, which are proportinal to the actual values of $\pi(\rho_j|y)$. $\pi(\rho|y)$ can be obtanied by fitting a curve (for example, using splines) to points $\{(\rho_j, \pi(y|\rho_j)\pi(\rho_j))\}_{i=1}^r$and re-scaling it to integrate one, similarly as INLA does.

Note that in our case, the values of the parameter of interest are bounded and this makes this approach easier. We may be interested in applying these 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 parametes is, say, λ, we could use an optimization algorithm to find the maximum (mode) of $\pi(y|\lambda)$. This will require 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 very computationally intersive and that it will not provide fast results, it should be noted that most computations can be done in parallel. In particular, the models that arise from diffrent values of λ can be fitted on separate nodes separate nodes of a cluster computer. We provide some computational times in Section 5 to show the feasibility of this approach.

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 a 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:

$$\pi(\rho, \gamma_l|y) \propto \pi(y|\rho, \gamma_l)\pi(\rho, \gamma_l)$$

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

$$\pi(\rho, \gamma_l) = \pi(\rho)\pi(\gamma_l)$$

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

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

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, it should be noted that this procedure can be easily paralellise to reduce computational time. Note also that accuracy can be improved by defining a thinner 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 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 the following Section.

4.1 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 will proceed as follows:

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

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

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

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 equaly spaced, then $\Delta_j = \Delta\ \forall j$.

Noting that

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

the approximation can be written as

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

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

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

Alternatively, this can be expressed as a weighted sums of the marginals distribution provided by each of the fitted models (for a given value of ρ):

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

Hence, marginal inference is still possible with this Bayesian Model Averaging approach (Hoeting et al., 1999). 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 paralellised to reduce computational time.

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

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

That is, the posterior mean of x_i is a weighted sum of the different posterior means computed conditioning on different values of ρ.

Note that in this case, if vague priors for the other parameters in the model, the parameter estimates should be very close to maximum likelihood estimates. In particular, if all the parameters in the model are assigned a flat (and probably improper) prior, this procedure will be equivalent to maximum likelihood.

4.2 Bayesian Model Selection

Alternatively, ρ can be regarded as a truly discrete variable and we may be interested in estimating the probabilities for each value. Hence, we will trying 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 of all, we will need to assign a prior distribution to each model. A vague prior is to give the same 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

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

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

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

so that

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

When a uniform prior is assigned to the models (i.e.,) the previous expression simplifies to

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

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 Bayesian Model Averaging. However, using a single model may produced biased inference on the parameters.

5 Simulation study

{sec:simstudy}

In order to assess the differences between the different models we have conducted a simulation analysis. In all cases, we are interested in how data created with one model are fitted by other models. In particular, we are interested in the degree of agreement between the classic Spatial Econometrics models and Generalizen all cases, we are interested in how data created with one model are fitted by other models. In particular, we are interested in the degree of agreement between the classic Spatial Econometrics models and Generalized Linear Models with random effects.

In order to simulate the datasets, we have considered the Gaussian and Binomial cases. In the Gaussian case, we have simulated 100 datasets considering a model with a single covariate and proper CAR random effect with variance-covariance matrix $\Sigma = \sigma^2(I - \rho W)$. This model can be written down as

$$y_i = \beta x_i + u_i;\ u_i \sim N(0, \sigma^2(I - \rho W)); i = 1, \ldots, 100$$

β has been set to 10, ρ to 0.5 and σ is 5. W is a row-standardised asjacency matrix. In order to develop the spatial structure, we will consider that observations are on a straight line, so that neighbors are the observations to the left and right. The edges are condiered as neighbours too.

Furthermore, we have simulated another 100 datasets according to each of the 3 spatial econometrics models described in Section 2.1. The parameters are the same as in the previous model. For the Spatial Durbin Model the coeffient of the lagged covariates has been set to 1.

Hence, for the Gaussian case we have 400 different data sets 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 very 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.

As measures of goodness-of-fit we will consider the following estimator for each model and group of simulated data sets:

$$ARMSE = \frac{1}{100}\sum_{k=1}^{100}\sum_{i=1}^{100}\frac{(y_i - \hat{y}_i)^2}{y_i}$$

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

5.1 Gaussian models

PRELIMINARY RESULTS:

All models seem to have a perfect fit to the data, i.e., in all cases the ARMSE for the 4 groups of datasets is 0!

5.2 Non-Gaussian models

PRELIMINARY RESULTS:

For the probit with CAR random effects the proporition of right classified is 97.14, 81.01, 83.52, 82.27. Hence, it provides almost perfect performance when applied to data generated under the same model and very good performance for data coming from spatial econometrics models.

6 Examples

{sec:examples}

6.1 Boston housing data

In our first example we will re-analyse the Boston housing data (Harrison and Rubinfeld, 1978). Here the interest is in estimating the median of owner-occupied houses using relevant covariates and the spatial structure of the data (Pace and Gilley, 1997). We 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. These results are summarised in Tables 1, 2 and 3.

	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

Table 1: Point estimates of the fixed effects for the Boston housing data using different models.

{tab:boston1}

Furthermore, we have displayed the posterior marginal distribution of ρ under different models in Figure 1. The different posterior distributions

	MLSDM	INLASDM
(Intercept)	1.898178	1.996513
CRIM	-0.005710	-0.005832
ZN	0.000691	0.000630
INDUS	-0.001113	-0.001495
CHAS1	-0.041632	-0.028907
I(NOX^2)	-0.010349	-0.006888
I(RM^2)	0.007950	0.007825
AGE	-0.001288	-0.001214
log(DIS)	-0.124041	-0.122187
log(RAD)	0.058635	0.055938
TAX	-0.000491	-0.000474
PTRATIO	-0.013199	-0.013531
B	0.000564	0.000540
log(LSTAT)	-0.247245	-0.249881
lag.CRIM	-0.004642	-0.004722
lag.ZN	-0.000379	-0.000257
lag.INDUS	0.000251	0.001275
lag.CHAS1	0.125183	0.100803
lag.I(NOX^2)	-0.386407	-0.418605
lag.I(RM^2)	-0.004513	-0.004207
lag.AGE	0.001497	0.001523
lag.log(DIS)	-0.004539	-0.011010
lag.log(RAD)	-0.009407	0.003412
lag.TAX	0.000411	0.000350
lag.PTRATIO	0.000603	-0.000588
lag.B	-0.000508	-0.000429
lag.log(LSTAT)	0.098467	0.091499

Table 2: Point estimates of the fixed effects for the Boston housing data using a Spatial Durbin Model. {tab:boston2}

should not be surprising as the models are in fact different and they have different spatial correlation structures.

6.2 After-Katrina business data

In this example we will look at the data analysed in LeSage et al. (2011) 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). Similarly as in the previous

	rho	rhosd	LLrho	ULrho
MLCAR	0.186	NA	NA	NA
INLACAR	0.970	0.0127	0.938	0.987
MLSEM	0.715	NA	NA	NA
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

Table 3: Estimates of the spatial autocorrelation parameter for the different models. {tab:bostonrho}

example, we have fitted four models. However, now we have used a GLM with a Binomial family and a probit link.

LeSage et al. (2011) split the data into four periods according to different time frames. In our analysis we will focus on the first period, i.e., the business re-opened during the first 3 months (90 days). The model used therein is the one that we have termed Spatial Lag Model in this paper.

7 Discussion and final remarks

{sec:discussion}

Bayesian spatial econometrics models play an importart role in the analysis of data with spatial structure. In this paper we have shown an alternative model fitting based on the Integrated Nested Laplace Approximation (INLA). Although it only provides marginal inference, INLA is computationally faster than MCMC. The authors have released the R-INLA package for the R programming language to fit a wide range of models. Although R-INLA software cannot fit the spatial econometrics models discussed in this paper it is possible to fit conditional models (by fixing one or more parameters in the full model) so that the final model can be obtanied by means of a Bayesian Model Averaging of the conditional models. 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 can be easily parallelised so that computing time is reduced.

We have compared standard spatial econometrics models and other models based on Generalized Linear Models with spatially correlated random effects. Our simulations show that model fitting using GLMs with random effects provides similar results to those (CHECK THIS!!!!).

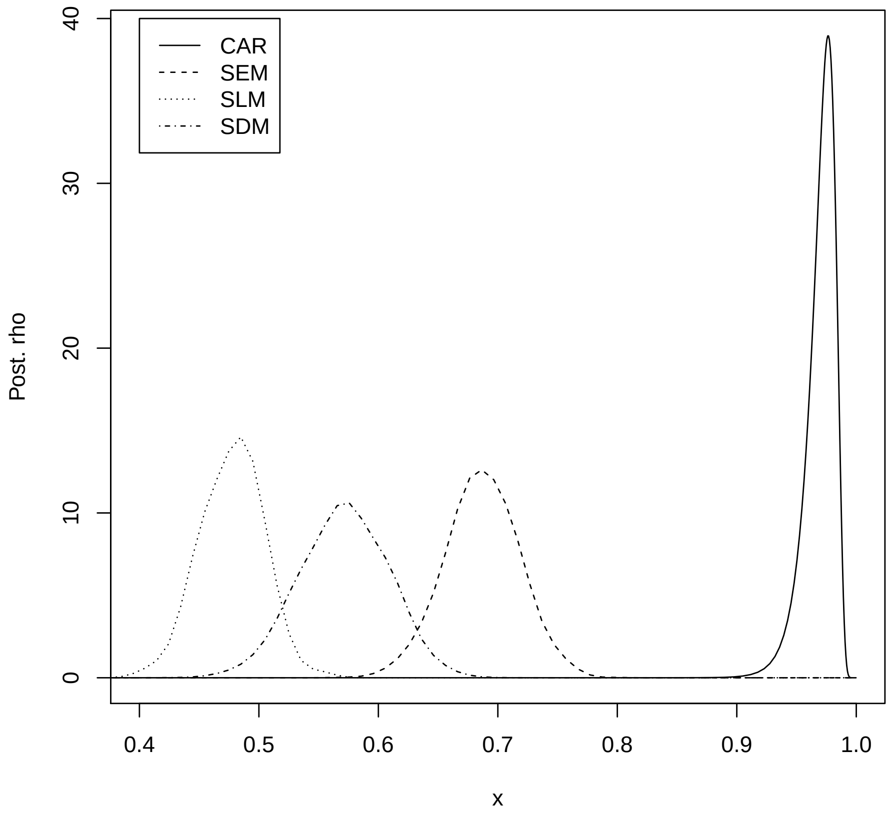

Figure 1: Marginals distribution for the spatial autocorrelation parameter ρ under different models. {fig:bostonrho}

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

References

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

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

	INLACAR	INLASEM	INLASLM	INLASDM
(Intercept)	-8.5136	-8.6076	-11.3811	-7.8558
flood	0.2691	0.3255	0.2862	0.2892
log(medinc)	0.9004	0.8356	1.1083	1.2858
sizeempsmall	-0.1318	-0.2718	-0.3082	-0.2881
sizeemplarge	-0.4292	-0.2828	-0.2933	-0.2982
sesstatuslow	-0.5072	-0.2534	-0.4243	-0.0949
sesstatushigh	0.0657	0.0699	0.0658	0.0807
owntypesole	0.3656	0.6063	0.6284	0.6382
owntypenational	0.1972	0.2098	0.1301	0.1155

Figure 2: Point estimates of the fixed effects for the Katrina data set using different models. {tab:katrina}

	rho	rhosd	LLrho	ULrho
INLACAR	0.5002	0.3738	0.002004	0.998
INLASEM	1.6094	0.7975	0.974572	0.990
INLASLM	0.0801	0.0636	0.001155	0.127
INLASDM	0.0918	0.0963	0.000846	0.117

Figure 3: Estimates of the spatial autocorrelation parameter for the different models on the Katrina dataset. {tab:katrinarho}

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.

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

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

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.

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.

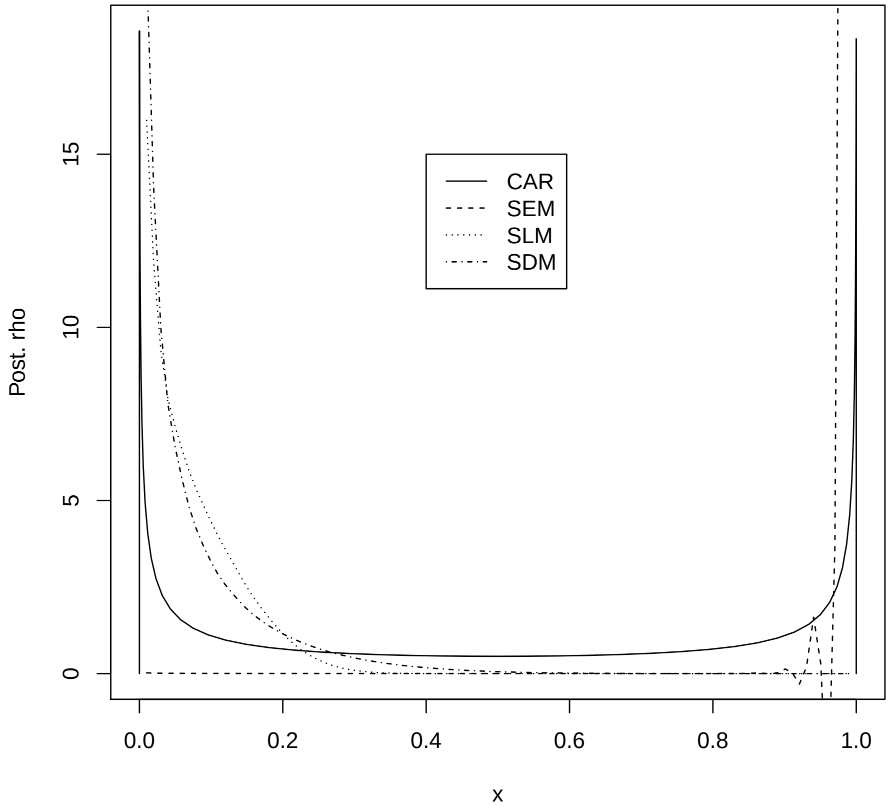

Figure 4: Marginals distribution for the spatial autocorrelation parameter ρ under different models. {fig:katrinarho}

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.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

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-STATISTICAL METHODOLOGY 71*(Part 2), 319–392.