Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 27 (2015) 116 – 118

Procedia Environmental Sciences

Spatial Statistics 2015: Emerging Patterns - Part 2

A new latent class to fit spatial econometrics models with Integrated Nested Laplace Approximations

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

[a] University of Castilla-La Mancha, 02071 Albacete, Spain
[b] Norwegian School of Economics, 5045 Bergen, Norway
[c] Adam Mickiewicz University, 61-680 Poznań, Poland
[d] Norwegian University for Science and Technology, 7491 Trondheim, Norway

Abstract

The new slm latent model for estimating spatial econometrics models using INLA has recently been introduced. It will be described briefly and its use will be demonstrated in the accompanying poster.

Keywords: Integrated Nested Laplace Approximation, Spatial econometrics

1. Introduction

Integrated Nested Laplace Approximation (INLA) provides a quick and reliable approximation to the marginal distributions in Bayesian inference on hierarchical models. This has been implemented in the R-INLA package which allows INLA to be used from within R statistical software. Although INLA is a general methodology, its use in practice is limited to the models implemented in the R-INLA package. In this poster, we describe the implementation and application of a new class of latent models in INLA made available through R-INLA as these models have not been available within R-INLA so far. This new latent class implements a standard spatial lag model that can be used to build more complex models in spatial econometrics. 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 poster. We will illustrate the use of this new latent model and its applications with two datasets based on Gaussian and binary outcomes.

2. The “slm” INLA latent model

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 [1]. We will follow the notation used in Bivand et al. [2], which is in turn derived from Anselin [3] and LeSage and Pace [4, 5].

*Corresponding author. Tel.: +47-55959355.
Email address: Roger.Bivand@nhh.no ()

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee
doi:10.1016/j.proenv.2015.07.119

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.

All these models can be rewritten so that the response y only appears on the left hand side. For example, the spatial error model (SEM) is:

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

and the spatial lag model (SLM):

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

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 (MCMC) is the most widely used. Rue et al. [6] 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. 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 θ_1. The vector $\mathbf{x}$ of latent effects forms a Gaussian Markov Random Field with precision matrix $Q(\theta_2)$, where θ_2 is a vector of hyperparameters. The hyperparameters can be represented in a unique vector $\theta = (\theta_1, \theta_2)$. It should be noted that observations $\mathbf{y}$ are independent given the values of the latent effects $\mathbf{x}$ and the hyperparameters θ. INLA will not try to estimate the joint distribution $\pi(\mathbf{x}, \theta|\mathbf{y})$ 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 different ranges depending on the number of latent variables and hyperparameters.

2.3. Using Bayesian model averaging to fit spatial economentrics models

Bivand et al. [2, 7] describe how to use Integrated Nested Laplace Approximation to fit some spatial econometrics models. Because of the lack of an implementation of these models within the R-INLA software at that time (the 2013 Spatial Statistics conference), they fit many different models conditioning on values of the spatial autocorrelation parameter. These conditioned models can be fitted with R-INLA and they are later combined using Bayesian model averaging (see Hoeting et al. [8]), to obtain the posterior marginals of the parameters of the desired model.

2.4. The "slm" latent model for fitting spatial economentrics models

The new "slm" latent model implements the following expression as a random effect that can be included in the linear predictor:

$$\mathbf{x} = (I_n - \rho W)^{-1} (X\beta + \varepsilon) \quad (3)$$

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, 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); logit(ρ) takes a Gaussian prior with zero mean and precision 10; and log(τ) takes a log-gamma prior with parameters 1 and $5 \cdot 10^{-5}$. Note that many spatial econometrics models can be derived from this implementation. In particular, the SEM model is a particular case with $\beta = 0$; the SLM model can be fitted with no modification. Details of the implementation are available at: http://www.math.ntnu.no/inla/r-inla.org/doc/latent/slm.pdf.

3. Examples

The examples shown on the poster are taken from the literature. First, we use the study by Harrison and Rubinfeld [9] of 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 (see Pace and Gilley [10]). The spatial adjacency that we will consider is for census tract contiguities. The second example data set covers business opening after Katrina; LeSage et al. [11] study the probabilities of reopening a business in New Orleans in the aftermath of hurricane Katrina. They have used a spatial probit, here we reproduce the analysis with a continuous link function (i.e., a probit function) and the new "slm" latent model.

4. Extensions

The work presented on the poster highlights only certain aspects of our results. Extensions include the calculation of impacts showing the consequences of feedback in models for which *beta* is included in slm models. They also include additional random effects, model selection and Bayesian model averaging in choices regarding the construction of spatial weights matrices.

References

[1] L. Anselin, Thirty years of spatial econometrics, Papers in Regional Science 89 (1) (2010) 3–25.
[2] R. S. Bivand, V. Gómez-Rubio, H. Rue, Approximate Bayesian inference for spatial econometrics models, Spatial Statistics 9 (2014) 146–165.
[3] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer, Dordrecht, 1988.
[4] J. LeSage, R. Pace, Introduction to Spatial Econometrics, CRC Press, Boca Raton, FL, 2009.
[5] J. P. LeSage, R. K. Pace, Spatial econometric models, in: M. Fischer, A. Getis (Eds.), Handbook of Applied Spatial Analysis, Springer, Heidelberg, 2010, pp. 355–376.
[6] H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, Journal of the Royal Statistical Society B 71 (Part 2) (2009) 319–392.
[7] R. S. Bivand, V. Gómez-Rubio, H. Rue, Spatial data analysis with R-INLA with some extensions, Journal of Statistical Software 63 (20) (2015) 1–31.
[8] J. Hoeting, A. R. David Madigan and, C. Volinsky, Bayesian model averaging: A tutorial, Statistical Science 14 (1999) 382–401.
[9] D. Harrison, D. L. Rubinfeld, Hedonic housing prices and the demand for clean air, Journal of Environmental Economics and Management 5 (1978) 81–102.
[10] R. K. Pace, O. Gilley, Using the spatial configuration of the data to improve estimation, Journal of the Real Estate Finance and Economics 14 (1997) 333–340.
[11] J. P. LeSage, R. Kelley Pace, N. Lam, R. Campanella, X. Liu, New Orleans business recovery in the aftermath of Hurricane Katrina, Journal of the Royal Statistical Society A 174 (4) (2011) 1007–1027.