Bayesian Selection of Spatio-temporal Autoregressive Models of Neighbourhood Effects: An Application to the Zaragoza Real Estate Market

A. Beamonte¹, P. Gargallo¹ and M. Salvador²

¹ Escuela Universitaria de Empresariales de Zaragoza
² Facultad de Ciencias Económicas y Empresariales de Zaragoza

Abstract: A Bayesian Selection of Spatio-temporal Autoregressive Models of Neighbourhood Effects (STAR) is proposed. MCMC algorithms are used in order to estimate the parameters of the models and to compare them. The methodology is applied to the analysis of the Zaragoza Real Estate Market.

Keywords: Real Estate; Bayesian Models Selection; STAR; Spatial-Statistics; MCMC.

1 Introduction

The evolution of the housing market is a subject of extraordinary interest from both academic and practical viewpoint. The importance of the Building sector in many economies owed to the number of jobs and the financial assets generated, as well as the high level of indebtedness of families due to the costly mortgages needed to buy a house, are some of the reasons for that interest.

From an economic and statistic point of view, the traditional model to analyze the value of a property is the hedonic pricing model proposed by Court (1939) and based theoretically by Rosen (1974). This model establishes that the market price of a property depends on its characteristics. In the particular case of the dwelling (Basu and Thiboudeau, 1998) this price is function of its structural characteristics (size, age, kind of construction, heating, lift, garage, storage room, air conditioner, number of rooms, floor, orientation, swimming-pool, etc.) and its location characteristics (distance to city centre, schools, kindergartens, underground stations, bus stops, supermarkets, level of air pollution, socioeconomic level in the neighbourhood, etc.)

The increasing capacities of the information systems and particularly, the geographic information systems (GIS) have made possible the construction of big databases with very precise spatio-temporal references at a microdatum level. In this context it would be necessary the implementation of
Bayesian Selection of STAR Models

Statistic techniques to build models describing the spatio-temporal dependences present in the data, in a correct and parsimoniously way. Likewise these techniques allows us to carry out more accurate inferences about the parameters and more exact predictions. There are two common difficulties in this kind of problems: on the one hand the high degree of non linearity of the parameters describing the spatial dependences and, on the other the huge sample sizes that make difficult the estimation of the model.

Recently, Pace et al. (1998, 2000) have proposed a family of autoregressive linear models that permits to capture an important number of local spatio-temporal effects. The technique is based on OLS procedures what make it low cost computationally. These authors apply their methodology to the price of houses obtaining better results than other complex techniques which use a much more number of parameters. Nevertheless, their model selection procedure is based on maximum likelihood estimation techniques whose asymptotic results are usually non-applicable (Stein (1999)). Moreover, in our opinion, this kind of models used to be less parsimonious and more difficult to interpret. In this paper we face these problems by means of a more parsimonious reparametrization of the model, using Bayesian techniques to estimate and select the models (Banerjee et al. (2003)) based on MCMC methods. The methodology is illustrated through a real data set from the Zaragoza Real Estate Market.

2 Preliminaries

Let $Y_{(n\times1)} \equiv (y_i; i = 1, \ldots, n)$ the vector of observations of the dependent variable where $y_i = \log(p_i)$ with $p_i$ the price of the $i^{th}$ transaction of a dwelling observed in a fixed period of time. The transactions have been ordered by date so $p_1$ is the oldest one, $p_2$ the second oldest, . . . , and $p_n$ the most recent. In addition, we consider $k$ exlanatory variables $X_1, \ldots, X_k$ related to the structural characteristics of the dwelling, location and neighbourhood externalities (we suppose $X_1 = 1$).

The model proposed to describe this dependence is a Spatio-temporal Autoregressive of Neighborhood Effects Model given by:

$$Y = X\beta + \phi_{TX}T(Y - X\beta) + \phi_{SX}S(Y - X\beta) + \phi_{TSX}TS(Y - X\beta) +$$

$$+ \phi_{STX}ST(Y - X\beta) + \phi_{TY} + \phi_{SY} + \phi_{STY} + \phi_{TSY} + \epsilon$$

where:

- $X_{(n\times k)}$ is the matrix of observations corresponding to the independent variables that have spatio-temporal effects, being $\beta_{(k\times 1)}$ the vector of regression coefficients which determine the effect type.

- $S_{(n\times n)}$ is the matrix that specifies the spatial relations between the transactions.
- \( T_{(n \times n)} \) is the matrix that specifies the temporal relations between the transactions.

- \(-1 < \phi_i < 1\) \(i \in \{TX, SX, TSX, STX, T, S, ST, TS\}\) are the autoregressive coefficients of the model which determine the effect transmitted by the characteristics and the prices of the previous transactions near in space and/or time.

- \( \epsilon \sim N_n(0, \sigma^2 I_n) \)

Prior distribution:

- \( \beta \sim N_k(0, \Sigma) \)
- \( \phi_i \sim U(-1, 1), i \in \{TX, SX, TSX, STX, T, S, ST, TS\} \)
- \( \tau = \frac{1}{\sigma^2} \sim Gamma\left(\frac{d_n}{2}, \frac{d_n \sigma^2}{2}\right) \)
- \( \lambda \sim U_D(\lambda_{\text{min}}, \lambda_{\text{max}}) \) with \(0 \leq \lambda_{\text{min}} < \lambda_{\text{max}} \leq 1\)
- \( m_S \sim U_D(m_{S,\text{min}}, m_{S,\text{max}}) \) with \(0 \leq m_{S,\text{min}} < m_{S,\text{max}} \in \mathbb{N}\)
- \( m_T \sim U_D(m_{T,\text{min}}, m_{T,\text{max}}) \) with \(0 \leq m_{T,\text{min}} < m_{T,\text{max}} \in \mathbb{N}\)

where \( m_S \) and \( m_T \) represent the number of nearest neighbours in space and time, respectively, \( U(a,b) \) denotes the uniform distribution in the interval \((a,b)\) and \( U_D(n, m) \) the discrete uniform distribution in \(n, n+1, \ldots, m\).

Given that the posterior distribution is analytically intractable, in this paper we develop an algorithm based on the Gibbs sampling in order to make inferences about the parameters of the model. The parameters are particularly relevant to determine the spatio-temporal dependences in the data. We will use Bayesian techniques for model comparison to select the more appropriate type of dependences.

### 3 Application to the Zaragoza Real Estate Market

In this section the model describe above will be apply to the analysis of the Zaragoza Real State Market. Tha data correspond to a sample of 788 transactions of dwellings carried out between November 2002 and December 2004 in the area pertaining to the Registro de la Propiedad number 5 of Zaragoza. The variables analyzed come from multiple administrative databases: Oficina Regional del Catastro, Instituto Aragonés de Estadística, Colegio de Registradores de la Propiedad and Ayuntamiento de Zaragoza. In the paper we describe the characteristics of the sample and show the results obtained in the model estimation and comparison process described in section 2.

**Acknowledgments:** This work has benefited from the financial support provided by Research Project 2006 0201 "Análisis Estadístico del precio de la vivienda en Zaragoza: una aproximación bayesiana" of the Fundación Economía Aragonesa-FUNDEAR
References


