Spatial Statistics 22 (2017) 47–67

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

MCMC Bayesian spatial filtering for hedonic models in real estate markets

P. Gargallo, J.A. Miguel, M.J. Salvador *

Departamento de Estructura e Historia Económica y Economía Pública, Facultad de Economía y Empresa, Universidad de Zaragoza, Spain

ARTICLE INFO

Article history:
Received 2 March 2016
Accepted 25 July 2017
Available online 14 August 2017

Keywords:
Bayesian inference
Variable selection
Adaptive MCMC
Hedonic models
Real estate market
Spatial filtering

ABSTRACT

The traditional hedonic model postulates that housing prices depend on their characteristics and their location. However, this model assumes a constant relationship between the dependent and the independent variables. This assumption is unrealistic because empirical studies have shown that the regression coefficients depend on the housing location. For this reason, it is necessary to use models with spatially varying coefficients. The approaches proposed in the literature used to estimate this type of models do not incorporate the uncertainty associated with the estimation and selection of models and/or are computationally expensive. To improve these aspects, this paper proposes spatial filtering techniques to parsimoniously model the spatial dependencies of the hedonic coefficients and an adaptive MCMC Bayesian algorithm to select the most appropriate filters. The method is illustrated through an application to the real estate market of Zaragoza, and a comparison with alternative procedures is conducted. Our results show that our valuation methodology has better goodness of fit and predictive performance properties than alternative methods. Although our proposal assumes normality and homoscedasticity of the model error distribution, the method is easy to implement and not very computationally demanding, which makes this approach attractive and useful from a practical viewpoint.

* Corresponding author.
E-mail addresses: pigarga@unizar.es (P. Gargallo), jamiguel@unizar.es (J.A. Miguel), salvador@unizar.es (M.J. Salvador).

http://dx.doi.org/10.1016/j.spasta.2017.07.010

1. Introduction

Real estate activities are linked to many sectors of the economy, including construction, finance, and insurance. Therefore, updated fair market housing values are extremely valuable to financial regulators and institutions, municipal assessors, housing index compilers, real estate developers, investors, and many others. Hedonic modelling is the most widely used method to estimate housing prices.

The traditional hedonic model postulates that housing prices mainly depend on their characteristics and locations. The model accounts for locational attributes (location of the dwelling and proximity to central business districts), neighbourhood attributes (availability of public schools, income levels and population density) and random spatial effects. However, it does not account for spatial interaction effects among dwellings. Empirical evidence shows that prices of neighbouring houses tend to be similar because they share common local factors, such as physical characteristics (age, size, and exterior and interior features) and amenities (socioeconomic status, access to employment opportunities, shopping, public service facilities, and schools). Moreover, information spillovers exist in housing markets, which are manifested through the spatial autocorrelation in prices (Wheeler et al., 2014). Besides, the hedonic model assumes a constant relationship between the dependent variable and the independent variables, which is an unrealistic assumption in housing markets, where it has been observed that the regression coefficients depend on housing location (Goodman and Thibodeau, 1998; Fotheringham et al., 2002; Páez et al., 2008; Wheeler et al., 2014).

Several reasons could explain the existence of relationship patterns that could be identified as market segments (Fotheringham et al., 2002; Wheeler et al., 2014). One reason relates to sampling variation because we would not expect the parameter estimates obtained from different samples to be the same. A second reason would be the spatial variations in the attitudes or preferences of people. For instance, the influence of the existence of a garage or a storage room on the price of a house is probably higher if the dwelling is located in the centre or the periphery of a city. A third reason could be the gross misspecification of the model due to omitted spatial explanatory variables or the assumption of an incorrect functional form. Hedonic theory provides little guidance on the choice of the functional form for the hedonic specification (Fleming, 1999).

To capture this spatial heterogeneity in housing markets, several modelling techniques have been proposed. Eckert (1990) suggested that, based on the assumption that subsets are characterized by a lower variance, models generated for housing submarkets should yield greater explanatory power (and predictive accuracy) than those computed at the overall market level. Goodman and Thibodeau (1998) introduced the concept of hierarchical linear modelling, whereby housing characteristics, neighbourhood characteristics, and submarkets interact to influence housing prices. Both of these approaches assume that the submarkets are previously known.

Brunsdon et al. (1996), Fotheringham et al. (1996, 2002) and Páez et al. (2002a, b) did not assume a previous knowledge of submarkets and proposed to estimate the regression coefficients for each dwelling using local geographically weighted regression (GWR) techniques. Using this method, an exploration of the variation of the parameters as well as a statistical analysis of the significance of this variation can be carried out. This methodology has received considerable attention in recent years, and some papers have applied GWR to housing markets (Brunsdon et al., 1999; Pavlov, 2000; Fotheringham et al., 2002; Yu, 2006 or Páez et al., 2008 among others). However, this method has been criticized because of the multicollinearity problems in the estimation of the parameters, which are due to the very similar characteristics of houses in the same area, which makes the estimation of the regression coefficients difficult (Wheeler and Tiefelsdorf, 2005; Griffith, 2008; Páez et al., 2011; Bárcena et al., 2014).

Several solutions have been proposed to address this problem. Wheeler (2007, 2009) and Bárcena et al. (2014) used penalized versions of GWR based on regularization methods (ridge and lasso regression) to build parsimonious models that weaken the multicollinearity problem and have good predictive and goodness of fit properties. Another alternative are the Bayesian spatially varying coefficients models (SVC) (Gelfand et al., 2003, 2004; Wheeler and Calder, 2007; Bárcena et al., 2014; Wheeler et al., 2014), which specify a single Bayesian hierarchical model that uses spatially varying coefficient processes to globally model the non-constant linear relationships between the variables.