Computers, Environment and Urban Systems 57 (2016) 1–11

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches

Marco Helbich [a,*], Daniel A. Griffith [b]

[a] Department of Human Geography and Spatial Planning, Utrecht University, Heidelberglaan 2, 3584 CS, Utrecht, The Netherlands
[b] School of Economic, Political and Policy Sciences, University of Texas at Dallas, 800 W. Campbell Rd, Richardson, TX 75080-3021, USA

ARTICLE INFO

Article history:
Received 4 March 2015
Received in revised form 23 October 2015
Accepted 8 December 2015
Available online xxxx

Keywords:
Hedonic models
Housing
Eigenvector spatial filtering
Spatial expansion method
Moving window regression
Geographically weighted regression
Prediction accuracy

ABSTRACT

Real estate policies in urban areas require the recognition of spatial heterogeneity in housing prices to account for local settings. In response to the growing number of spatially varying coefficient models in housing applications, this study evaluated four models in terms of their spatial patterns of local parameter estimates, multicollinearity between local coefficients, and their predictive accuracy, utilizing housing data for the metropolitan area of Vienna (Austria). The comparison covered the spatial expansion method (SEM), moving window regression (MWR), geographically weighted regression (GWR), and genetic algorithm-based eigenvector spatial filtering (ESF), an approach that had not previously been employed in real estate research. The results highlight the following strengths and limitations of each method: 1) In contrast to SEM, MWR, and GWR, ESF depicts more localized patterns of the parameter estimates and does not smooth local particularities. 2) ESF is less affected by multicollinearity between the local parameter estimates than MWR, GWR, and SEM. 3) Even though the in-sample explanatory power and prediction accuracy of ESF is superior compared to the competitors, repeated sampling indicates a limited out-of-sample fit and prediction accuracy, suggesting over-fitting tendencies. 4) The application of ESF is less intuitive than MWR and GWR, which are available off-the-shelf.

1. Introduction

Interest in hedonic models that consider the spatial heterogeneity of pricing effects to explore real estate markets in urban areas has grown rapidly (Helbich, Brunauer, Hagenauer, & Leitner, 2013; Lu, Charlton, Harris, & Fotheringham, 2014). Conventional global hedonic models assume a unitary housing market across space that can be modeled through a single price function being representative throughout a city (Bitter, Mulligan, & Dall'erba, 2007). Such models are increasingly questioned due to their unrealistic simplification of housing markets (McMillen & Redfearn, 2010). As a consequence, local hedonic models emerged as an alternative to explore spatially varying housing prices. Even though spatially varying pricing effects are congruent with urban economic theory (Redfearn, 2009) referring to "micro-market effects" (Sunding & Swoboda, 2010, p. 558), and emerging where local legislation and policy regulation are effective (Helbich, Brunauer, Vaz, & Nijkamp, 2014), their incorporation in hedonic models constitutes a methodological challenge. However, neglecting spatial heterogeneity might have serious consequences for model estimation, such as biased regression coefficients, resulting in inappropriate conclusions (LeSage & Pace, 2009; Páez, Fei, & Farber, 2008). No less important, since policy strategies rely on such models, it is critical for decision makers to have models that have the highest fit (Ahn, Byun, Oh, & Kim, 2012; Bourassa, Cantoni, & Hoesli, 2010) and that inform them properly about local housing market conditions, for example through visualizations of spatially varying marginal prices (Ali, Partridge, & Olfert, 2007). Such models also reduce the risk for mortgage lenders and appraisal agencies by obviating loan losses and erroneous real estate assessments.

Despite these appealing methodological and practical advantages of localized models (e.g. Fotheringham, Charlton, & Brunsdon, 2002; Griffith, 2008) in real estate applications, there is still disagreement over which local hedonic approach is superior (Ahn et al., 2012). In this regard, comparative studies are helpful to contrast the merits of different modeling techniques, particularly in light of the increase in the number of applications and the proliferation of new approaches (Páez et al., 2008). Until now, simulation experiments based on artificially generated data with known properties have dominated the comparative analysis literature (e.g., Páez, Farber, & Wheeler, 2011). Even though such investigations greatly improve our knowledge of the advantages and limitations of specific hedonic models, without linking them to more complex real-world case studies, simulation studies cannot entirely uncover their practical relevance. Consequently, empirical model assessments complementing simulations are essential. As model competition outcomes are data-dependent and might cause contradictory results, Bourassa et al. (2010) recommend that empirical comparisons utilize a single dataset.

* Corresponding author.
E-mail address: m.helbich@uu.nl (M. Helbich).

http://dx.doi.org/10.1016/j.compenvurbsys.2015.12.002
0198-9715/

Therefore, the principal objective of this study was to address the model performance of four spatially varying coefficient models using a housing dataset for the metropolitan area of Vienna, Austria. As opposed to Farber and Yeates (2006) and Bitter et al. (2007), this study compared SEM, MWR, and GWR by applying a more rigorous out-of-sample accuracy assessment, resulting in less optimistic results than when using the R^2 as performance measure. There are three reasons for selecting these models: their performance is good, they have remarkable recognition in urban housing studies, and they support an enhanced understanding of local market conditions (e.g. Helbich et al., 2014; Kestens, Theriault, & Des Rosiers, 2006; Osland, 2010; Sunding & Swoboda, 2010). The second innovation was the introduction of ESF to model geographically varying relationships and to test the predictive performance of this approach relative to SEM, MWR, and GWR. It is this model, which had not previously been utilized in the context of hedonic modeling, that makes this study not only of interest for urban analysis, but also of practical relevance to urban policymaking. Finally, as ESF grounds on stepwise variable selection procedures which only test a limited number of variable combinations (i.e., the interaction terms between the eigenvectors and housing predictors), a genetic algorithm-based approach had been proposed as alternative.

2. Spatial hedonic price analyses

The theoretical foundation of hedonic modeling is motivated by Lancaster's (1966) theory of consumer utility, which argues that it is not the good itself that generates utility, but the good's specific characteristics. Grounded in this notion, Rosen (1974) developed hedonic pricing theory, which explains that a house price is the sum of its utility-bearing characteristics. Housing is thus considered a heterogeneous good consisting of non-separable structural and neighborhood features (Malpezzi, 2003). Each of these characteristics has its individual implicit price. Because property is fixed in space, a household implicitly chooses a bundle of different goods by selecting a specific house, seeking to maximize its utility. Hence, a household's purchasing decision theoretically reflects an optimal configuration of housing attributes and their paid transaction price (Sheppard, 1997).

Hedonic analysis provides a well-established approach to deconstruct a total house price, and to determine corresponding marginal prices (Malpezzi, 2003). A hedonic equation and its associated unknown parameters are estimated through non-spatial and spatial econometric regression or geostatistical approaches (e.g., Anselin & Arribas-Bel, 2013; Kuntz & Helbich, 2014). Besides the specification of the functional form (Helbich, Jochem, Mücke, & Höfle, 2013), spatial effects subsuming spatial autocorrelation (SAC) and spatial heterogeneity, challenge model estimation (Dubin, 1998). Spatial effects are deduced from the durability and spatial fixation of properties, questioning the validity of non-spatial regression (McMillen & Redfearn, 2010). Accordingly, by assuming spatial equilibrium between supply and demand, one global regression model is assumed to be valid for an entire market, and the estimated parameters are constant across space. Once a dwelling is constructed, it becomes immovable, and supply becomes inelastic (Schnare & Struyk, 1976). These supply inelasticities are coupled with a differentiation in demand emerging from dissimilar households (e.g., due to income variation, diverse socioeconomic characteristics), which value housing properties differently (Quigley, 1985). Both issues cause local supply–demand imbalance (Bitter et al., 2007) and challenge unitary housing markets. Therefore, functional disequilibrium and housing market segmentations are rational (Goodman & Thibodeau, 2003; Kestens et al., 2006), causing distinct patterns of price differentials that manifest as spatially heterogeneous marginal prices (Palm, 1978). Consequently, if this assumption of market segmentation is accepted, but not appropriately modeled, the hedonic coefficients are biased and models have a loss of explanatory power (Bitter et al., 2007; Bourassa et al., 2010; Helbich et al., 2014; Schnare & Struyk, 1976), while local price variations remain hidden.

3. Modeling spatial variation: a review

Spatially varying coefficient models emerged to circumvent the limitations of using spatial regimes in global models, for example, that discrete market boundaries are known in advance and homogeneity within each region is present (Anselin & Arribas-Bel, 2013). Since spatial regimes were not relevant to the present study, the subsequent sections deal only with SEM, MWR, and GWR.

3.1. Spatial expansion method

A classic approach to model spatial structural instability is Cassetti's (1972, 1997) SEM (see Section 4.2), a precursor of GWR. Here, global coefficients are parameterized by polynomials, where covariates are expanded by spatially explicit variables within an ordinary least squares (OLS) framework (Fotheringham, Charlton, & Brunsdon, 1998). However, Pace, Barry, and Sirmans (1998) showed that a polynomial expansion is too imprecise to model spatial variation effectively. While polynomials have appealing usage, they lack robustness and tend to over-smooth local variation, and higher-order polynomials induce multicollinearity. Nevertheless, SEM has received attention in the real estate context from Can (1992), Kestens et al. (2006), and Bitter et al. (2007). For instance, Can (1992) interacted a small set of structural housing variables with neighborhood quality to model spatial drifts. Complementing Can (1992), Fik, Ling, and Mulligan (2003) utilized a fully interactive model that includes higher-order polynomials. Due to numerous interaction terms, Fik et al. (2003) had to limit the number of structural characteristics. Because such a reductionistic model is affected by omitted variables, its estimates are most likely biased. Although SEM is an improvement over global models (Pavlov, 2000), it is criticized for its inability to capture spatial trends other than those that are non-complex and broad, simultaneously discarding valuable local variation. In contrast to Pavlov (2000), who relaxed the parametric assumption of SEM by using non-parametric functions of spatial coordinates, Fotheringham et al. (2002) promoted moving window approaches.

3.2. Moving window and geographically weighted regression

Both MWR and GWR (Fotheringham et al., 2002) circumvent the modeling inflexibility problems of SEM. GWR extends MWR through additional distance-based weightings (see Section 4.3). A benefit of MWR and GWR is that marginal prices are allowed to vary smoothly across space by setting regional dummies or polynomial expansions aside. From a theoretical viewpoint, Bitter et al. (2007) argue that, by restricting the number of sales per local regression, GWR partly mimics appraisers' sales comparisons and price adjustment processes. Despite these appealing properties, GWR is under debate. For example, Wheeler and Tiefelsdorf (2005) and Griffith (2008) referred to multicollinearity problems amongst GWR estimates. While weak correlation affects the ability to interpret model output, strong dependencies make a reliable separation of individual variable effects hardly possible (Wheeler & Tiefelsdorf, 2005). Páez et al. (2011) noted that GWR itself artificially introduces multicollinearity, even if the input covariates are uncorrelated, while Jetz, Rahbek, and Lichstein (2005) reported sign reversals that can be traced back to multicollinearity, causing a local omitting variable bias. However, model calibration, which is based on predictive performance, remains unaffected (Brunsdon, Charlton, & Harris, 2012). Others, including Wheeler (2009) and Vidaurre, Bielza, and Larrañaga (2012), have proposed integrating ridge and lasso regression into GWR to alleviate collinearity complications (Ahn et al., 2012). However, these extensions have not found resonance in real estate. Fotheringham et al. (2002) examined the calibration procedures of hedonic GWR models and concluded