Land Use Policy 79 (2018) 116–122

Contents lists available at ScienceDirect

Land Use Policy

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

Large-scale simultaneous market segment definition and mass appraisal using decision tree learning for fiscal purposes

Fabián Reyes-Bueno, Juan Manuel García-Samaniego, Aminael Sánchez-Rodríguez*

Departamento de Ciencias Biológicas, Universidad Técnica Particular de Loja.,San Cayetano Alto s/n, PC: 110104, Loja, Ecuador

A B S T R A C T

Cadastral assessment aims at guarantying equity in the allocation of property taxes. Therefore, we must be able to massively determine property values through models that reflect, with the minimum error, the behaviour of land market in each region. Despite this imperative need, currently land valuation for cadastral purposes is plagued with subjectivity. A very extended bad practice for instance is to assume that variables of productive performance i.e. land use capacity, are the ones with the highest influence on land value formation in the rural sector. The former assumption largely ignores the plethora of rural land uses that exist nowadays. To open the door to less subjective methodologies of land mass appraisal we borrowed statistical methodologies from the field of data-mining and applied them to a dataset of 410 purchase-sale transactions (2003–2009) of land plots located in the rural sector of the Vilcabamba parish (southern Ecuador). Land market behaviour in Vilcabamba responds to a transition from a pure agricultural territory to a touristic one at which many second-homes are being built for leisure. Our results demonstrate the applicability of methodologies such as model-tress (M5P) and multivariate adaptive regression splines (MARS) to rural land mass appraisal. Both M5P and MARS allow defining market segments while simultaneously establishing the weights of predictor variables for land value formation. We also collected evidence supporting that removing variables of productive performance from land value prediction models do not hamper models predictive power at least in rural areas where gentrification is taking place.

1. Background

Cadastral assessment is a mass appraisal process of property groups used for calculating the real property tax (Baumane, 2010). During cadastral assessment a property value is commonly calculated by value determination models which aim at reducing errors during value estimation. Cadastral assessment is very important to ensure right real property taxation and the principles of equality (Baumane, 2010). In Ecuador, the cadastral value of a given property is the basis for taxation and for establishing rates (*e.g.* the corresponding percentage for fire brigades) and special contributions (*e.g.* to pay for a specific infrastructure work on a given area). Cadastral value of a property in Ecuador is also taking into account expropriation and compensation processes. According to the Ecuadorian legislation (Ecuador, 2010) the cadastral value of a property in a given sector should be established by adding to the land value, the property value itself. Property values are determined by comparing against unit prices of comparable properties from the same sector. The resulting cadastral (market) value is then the most probable price (in terms of money) of a property in a competitive and open market provided the conditions needed to guarantee a fair sale. Both the buyer and seller must act prudently and knowledgeably, and it is commonly assumed that the price is not affected by undue stimulus (Iaao, 2011).

From the methodological point of view, cadastral assessment consists of three stages (Fig. 1):

1.1. Stage 1: seed points selection

A prerequisite for cadastral value formation is the identification of comparable properties in the same sector to which the property being assessed belongs. To this end its important to possess enough land market information from the whole study site. Once enough market information is available, one can proceed to define a series of “seed points” (georrefered) around which market values are expected to behave homogeneously.

1.2. Stage 2: homogeneous zones definition

From the “seed points” identified during Stage 1, homogeneous zones (HZs) are defined: zones in which land market is expected to behave homogeneously. Market homogeneity in this context means that within a HZ, the coefficient that modify the values taken by variables that have an effect on land value formation remain constant. It could be said then, that HZs definition is no more than finding the geographical areas where coefficients affecting the variables that best explain value formation are truly constant. In this sense, HZs are also known as

* Corresponding author.
E-mail addresses: frreyes@utpl.edu.ec (F. Reyes-Bueno), mgarcia@utpl.edu.ec (J.M. García-Samaniego), asanchez2@utpl.edu.ec (A. Sánchez-Rodríguez).

https://doi.org/10.1016/j.landusepol.2018.08.012
Received 4 April 2018; Received in revised form 6 August 2018; Accepted 6 August 2018
0264-8377/ © 2018 Elsevier Ltd. All rights reserved.

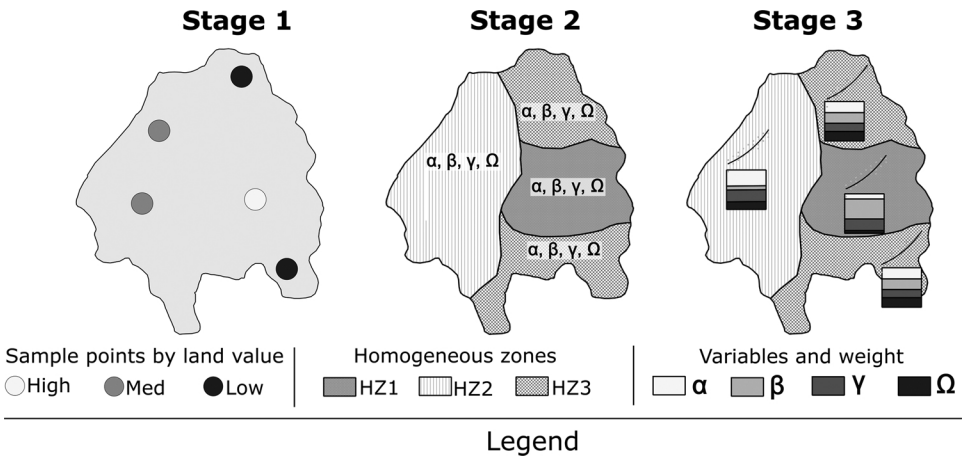

Fig. 1. The process of land value predicting models generation. **Stage 1**: the territory under study is sampled at points of known land value (light grey circles represent high land values, grey circles represent average land values and black circles represent low land values). **Stage 2**: as an example, four homogeneous zones (HZs) are identified (solid black lines) according to the behaviour of spatial variables that most influence land value formation. **Stage 3**: Land formation models are generated and the relative weight of each variable in the resulting equation determined (represented as the height of horizontal bars) within each HZ.

market segments or submarkets. When they are properly defined, HZs can be instrumental for estimating land values and for prioritizing the most important variables for value formation at each HZ (Lozano-Gracia and Anselin, 2012).

Currently in Ecuador, HZs are being defined in a complete subjective fashion. In the rural sector for instance, HZs definition is based on variables related to land use capacity such as slope, soil texture, effective depth, stoniness and drainage. Other variables commonly used for HZs definition in rural Ecuador are climatic ones e.g. precipitation, hydric deficit and temperature (Dinac et al., 1989; Magap, 2008) which are most of the time employed with redundancy. To date, there is no analysis that effectively demonstrates a decrease in heterogeneity during land value estimation thanks to the use of land use capacity or climatic variables. In the present work, we will use a plethora of mathematical approaches for HZs definition in an unsupervised fashion i.e. data-driven detection. The use of unsupervised methods is done to avoid any source of subjectivity during HZs definition i.e. prioritizing certain variables over the other (see further).

1.3. Stage 3: land value predicting models generation

It is important to note that at each HZ, land has a "base value" which is determined by the magnitude of several coefficients, each affecting a variable that resulted important for value formation (the set of prioritized variables at each HZ during Stage 2). In a perfectly modelled land market, the land base value should change from one HZ to the other as the coefficients (and the variables) change reflecting a locally adapted model. However, in the Ecuadorian context, the determination of the base value associated with a given HZ is so unreal that the weight assigned to the value-forming variables is the same among all the zones identified in a territory (Magap, 2008).

There is no universally accepted method or technique for the identification of HZs (their surface, limits, etc.) and to model the process of land value formation at each of the HZs (Kennedy et al., 1997). Among the techniques that have been used to this end, are: the geographic weighted regression (Hayles, 2006; Manganelli et al., 2014), the cluster analysis (Hayles, 2006; Kennedy et al., 1997), the main component analysis (Kennedy et al., 1997), and CART decision tree (Valenti et al., 2015).

By the end of Stage 3, we should obtain a model for the prediction of the land value. Such model is mainly based on the weights assigned to the coefficients as to modify the value-forming variables in a way that best reproduce the value changes that occur among all HZs in a given territory. An ideal model should on the one hand comprise only the variables that best explain the variance in the input data. On the other hand, it should give information on the weight each of these variables has on land value formation, which could in fact be different across HZ. One of the most widely used techniques for obtaining land value models is the multiple regression analysis (MRA) (Buurman, 2003; Elad et al., 1994; Hayles, 2006), although its application at large scales e.g. a canton, is not appropriate (Kauko and d'Amato, 2009; Mora-Esperanza, 2004). The former is because MRA is not able to properly capture the spatial (non-linear) dependence that the land value has on the land market dynamics that occur in a large territory.

As we have seen so far, the application of methodologies such as MRA for the generation of land value predicting models are based on the pre-definition of the HZs from which a final model is constructed. However, there are alternative techniques that could simultaneously segment the market and model the relationships of the variables impacting value formation (combining stage 2 and 3 into a single process). The main advantage of simultaneously defining HZs and variable weights is that the models (one per each HZ) are generated from the input data as a whole which maximizes the amount of variation finally explained by the models. That is, HZs definition becomes a data-driven process within a single dataset. Resulting models could then be used to identify the influence exerted by the explanatory variables on land value in each homogeneous zone (Clifton and Spurlock, 1983).

However, the techniques that would allow the simultaneous execution of Stages 2 and 3 have been poorly used in the field of cadastral assessment in a systematic way. Among such techniques, decision trees (DT), which are very advantageous for land markets, allow the model to adapt to local characteristics that condition it. There are several DT algorithms, including Model Tree (MT) and Multivariate Adaptive Regression Splines (MARS), which generate subsets of values showing small variations among them and generates regression functions for each subset (Wang and Witten, 1996). These techniques can face problems of classification and regression, are easy to interpret, and are of great help to analyze linear and nonlinear relationships between the dependent variable and the independent ones (Fan et al., 2006). The Model Tree technique was applied by (Acciani et al., 2008) to model the price of 109 vineyard properties in Southern Italy, obtaining more satisfactory results than with MRA on the same dataset.

Taking into account how little explored the DT method has been for cadastral valuation, in the present study we seek to answer the following questions: is the DT method suitable for the rural cadastral valuation? Can the variable land capacity improve the predictive power of land value predicting models? Have the variables affecting value formation the same weight across HZs? To answer such questions, the present study was carried out in the parish of Vilcabamba (Loja province, Ecuador). Due to the accelerated process of land transfer that has experienced Vilcabamba (Reyes-Bueno et al., 2016), we were able to have enough samples of rural properties to model land market in this territory. Four techniques for the generation of land value predicting models were compared: Linear regression, M5P model tree, M5P model tree with Bagging, and Multivariate Adaptive Regression Splines - MARS. The results show that model trees outperform all other methods

Download English Version:

https://daneshyari.com/en/article/6545982

Download Persian Version:

https://daneshyari.com/article/6545982

Daneshyari.com