European Journal of Operational Research 217 (2012) 120–128

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Valuating residential real estate using parametric programming

Subhash C. Narula [a], John F. Wellington [b,*], Stephen A. Lewis [b]

[a] *Virginia Commonwealth University, Richmond, VA 23284-4000, USA*

[b] *Indiana University – Purdue University Fort Wayne (IPFW), Fort Wayne, IN 46805-1499, USA*

ARTICLE INFO

Article history:
Received 6 January 2010
Accepted 14 August 2011
Available online 24 August 2011

Keywords:
Linear programming
Parametric programming
Real estate valuation
Regression
Regression quantiles

ABSTRACT

When the estimation of the single equation multiple linear regression model is looked upon as an optimization problem, we show how the principles and methods of optimization can assist the analyst in finding an attractive prediction model. We illustrate this with the estimation of a linear prediction model for valuating residential property using regression quantiles. We make use of the linear parametric programming formulation to obtain the family of regression quantile models associated with a data set. We use the principle of dominance to reduce the number of models for consideration in the search for the most preferred property valuation model (s). We also provide useful displays that assist the analyst and the decision maker in selecting the final model (s). The approach is an interface between data analysis and operations research.

1. Introduction

The objective of this paper is the presentation of a meaningful method for valuating single-family residential property using a hedonic model that incorporates features of the property such as its age, square feet of living space, lot size, number of rooms, and others. The underlying thesis of the hedonic model is that the valuation of the residence can be related to a 'bundle' of the property's features (Kummerow, 2000). This principle is used in valuating residential property for "purchase and sale, transfer, tax assessment, expropriation, inheritance or estate settlement, investment and financing … by real estate agents, appraisers, mortgage lenders, brokers, property developers, investors and fund managers, lenders, market researchers and analysts, shopping center owners and operators, and other specialists and consultants" using multiple linear regression methods (Pagourtzi et al., 2003). Although modeling residential property value in this manner is not the only technique, regression methods are commonly and routinely used in mass appraisal and other areas of real estate (Ferreira and Sirmans, 1988). In fact, according to the literature, "Appraisers must supplement their skill set with valuation methods that can systematically analyze larger data sets with output that is readily applicable to single-property appraisal. The importance of this cannot be overstated. These systems use statistical models to derive real estate value, replacing flesh and blood appraisers. They also use all available market data, most often in the form of a database of comparable sales," (Kane et al., 2004). They continued: "Appraisal valuation modeling techniques augment traditional appraisal practice. The appraiser, therefore, is maintained as the valuation expert." This point is particularly important in that the method proposed in this paper positions the valuation expert centrally in selecting the final valuation model.

In this paper, the single equation hedonic linear regression model is used to valuate residential property using the method of quantile regression (QR) due to Koenker and Bassett (1978). QR has very appealing aspects that translate well to valuating residential property. It is very descriptive and offers a focus on the changes (regression residuals) in property valuations produced by the models. The latter is particularly meaningful because it is the source of satisfaction and otherwise for parties directly impacted by the valuation such as property owners and taxing authorities. We refer to this as the loss associated with changes in property valuation. Because QR produces many regression models, it provides the analyst and decision maker with alternate models to consider in controlling loss arising with model implementation. When residential property is valuated above a threshold percent that reflects the owner's perception of its fair valuation, the owner may challenge the new valuation. However, the owner may not do so if the new valuation is less than the current valuation. At the same time, property valuations are intended to produce tax revenue. Therefore, it is desirable to find a valuation model that is fair to the tax authority and to property owners. The tax authority should not lose tax revenues and properties should not be unduly over-valuated. We find that quantile regression is well suited to incorporating these implementation concerns. We

* Corresponding author. Tel.: +1 260 481 6061; fax: +1 513 481 6879.
E-mail addresses: snarula@vcu.edu (S.C. Narula), wellingj@ipfw.edu (J.F. Wellington), lewis.sa07@gmail.com (S.A. Lewis).

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.08.014

note that challenges to new property valuations are expensive to resolve.

The intent of the paper is to illustrate the utility of valuating residential property using the hedonic linear regression model and parametric programming. The focus is on the loss resulting from model implementation and not the statistical precision of the estimated regression coefficients or the performance of the hedonic model vis-a-vis other specifications of residential property valuation. The valuation techniques addressed in this paper are comparative methods that valuate property in the company of other properties that share a common feature such as location or a temporal aspect such as members of a set of properties scheduled for periodic re-valuation.

The rest of the paper is organized as follows. In the next section, we review regression modeling of residential property valuation under various criteria including regression quantiles and provide an example. In Section 3, we present a brief literature review of methods for valuating residential property and regression modeling of the same with emphasis on quantile regression. The mathematical parametric programming formulation of the quantile regression problem is given in Section 4 and discussion of model selection appears in Section 5. We conclude the paper with remarks in Section 6.

2. Regression modeling of residential property valuation

For a single equation multiple linear regression model, let y denote the $n \times 1$ vector of observed values of the response variable corresponding to X, the $n \times k$ matrix of the values of k predictor (or regressor) variables that may include a column of ones to represent an intercept term. Then

$$y = X\beta + \varepsilon \tag{1}$$

where β is the $k \times 1$ vector of unknown parameters and ε is the $n \times 1$ vector of unobservable random disturbances in y. In the application of (1) to valuating residential property, y represents the current valuations of single-family residential properties; X, the physical characteristics or attributes of the properties; and n, the number of properties to be valuated.

When the single equation linear regression model (1) is used for property valuation, the regression residual is the magnitude of the adjustment in the property's valuation. The negative residual indicates that the valuation obtained from the regression model is above the current valuation and increases the tax base and tax revenue derived from it. The positive regression residual indicates the contrary. When the property is valuated above (below) a threshold percent of perceived fair adjustment, the owner may (not) challenge the new valuation. Hence the loss (change in tax base and the number of challenges to new property assessments) associated with implementing valuations derived from the regression model are related to the absolute and relative magnitudes of the regression residuals. The net increase in property valuations is the sum of the absolute negative regression residuals minus the sum of positive residuals.

Consider the real estate data (available at http://users.ipfw.edu/wellingj/) that consists of 54 observations on y, the current valuations of the set of properties, and ten predictor variables $x_1, \ldots, x_{10}$ that represent respectively taxes, number of baths, frontage (feet), lot size (square feet), living space (square feet), number of garages, number of rooms, number of bedrooms, age of home (years), and number of fireplaces. Because y is zero when the values of variables $x_1, \ldots, x_{10}$ are zero, the intercept term is omitted in modeling the data in the manner of (1).

2.1. Least squares, minimum sum of absolute errors, and multiple criteria regression models

The least squares (LS) regression modeling of the data resulted in net increase in property valuations of −$8,545, i.e. if the model were used to valuate the properties, the tax base for the fifty-four properties would be $8,545 below current aggregate valuations, see Table 1. Fitting the data to (1) under the minimum sum of absolute errors (MSAE) criterion produced net increase of $155,496. The maximum relative increase in valuation is 45.89% for the LS model and 67.18% for the MSAE model. For the LS result, the number of valuations that would increase by at least 10% and 20% is 16 and 7, respectively; for the MSAE model, the counts are 14 and 6 respectively.

Narula and Wellington (2007) proposed a multiple criteria methodology for valuating residential properties. The results of maximizing the net increase in property valuations subject to five bounds ($\leqslant$60%, 50%, 40%, 32.5% and 31.5%) of allowable relative change in any property valuation are reported in Table 1. The net increase in property valuations for Models 1–3 exceeds the values for the LS and MSAE models. However, the number of property valuations above 10% and 20% of their current values for each of the five models is higher than the counts for the LS or the MSAE models.

2.2. Quantile regression and parametric programming

Koenker and Bassett (1978) formulated the regression quantile problem as a linear parametric programming problem and as such defined a family of regression models. The formulation is a function of a single parameter that describes the fraction of the regression residuals with negative values. The parameter is often denoted by θ and defined over the interval [0, 1]. When applied to valuating residential property, the parameter describes the fraction of property valuations in the data set that are valuated above current values (y). The number of regression quantile models associated with a data set is of order n.

When the value of θ equals zero, all regression residuals are non-negative, i.e., all properties valuations derived from the $\theta = 0$ regression quantile model are no greater than current values. On the other hand, when the value of θ equals one, all residuals are non-positive, i.e., all property valuations obtained from the $\theta = 1$ regression quantile model are at or above current values. Clearly, the regression quantile models for θ near zero are not desirable because the resulting tax base would be smaller and the tax authority would lose revenue; for θ near one, many of the resulting valuations may be above the property owners' perceived thresholds of fair adjustment and in consequence generate many challenges by property owners.

Fig. 1 is the display of the empirical regression quantile function (net increase in property valuations versus θ) for the real estate

Table 1
The loss measures for the LS, MSAE, and multiple criteria regression models.

Model	Maximum percentage change in valuations	Net gain in valuations ($000)	No. of valuations increased 10% or more	No. of valuations increased 20% or more
LS	45.89	−8.545	16	7
MSAE	67.18	155.496	14	6
Multiple criteria models				
1	60	2579.395	39	32
2	50	1776.410	35	26
3	40	865.262	30	16
4	32.5	54.455	17	10
5	31.5	−61.362	17	10