VALUATING RESIDENTIAL REAL ESTATE USING PARAMETRIC PROGRAMMING

Subhash C. Narula
Virginia Commonwealth University
Richmond, Virginia 23284-4000 USA
E-mail: snarula@vcu.edu

John F. Wellington
Richard T. Doermer School of Business
Indiana University – Purdue University Fort Wayne (IPFW)
Fort Wayne, IN 46805-1499 USA
E-mail: wellingj@ipfw.edu

Stephen A. Lewis
Richard T. Doermer School of Business
Indiana University – Purdue University Fort Wayne (IPFW)
Fort Wayne, IN 46805-1499 USA
E-mail: lewis.sa07@gmail.com

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.

Keywords: Linear programming; parametric programming; real estate valuation; regression; regression quantiles.

---

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 multiple 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 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 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 nx1 vector of observed values of the response variable corresponding to X, the nxk 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 kx1 vector of unknown parameters and ε is the nx1 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, …, 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, respectively. 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 results, 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 (≤ 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 at or below 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.

Figure 1 is the display of the empirical regression quantile function (net increase in property valuations versus θ**)** for the real estate data. In Table A.1 of the Appendix for each regression quantile model, we noted the associated: i) maximum percentage increase in valuation; ii) net increase in valuations, i.e., the sum of the increases in property valuations minus the sum of the reductions produced by the model; iii) the number of valuations increased ten percent or more above current values; and iv) those increased twenty percent or more. Changes (regression residuals) in the property valuations for the few select regression quantile models addressed in this paper are displayed in Figure 2 and Figure 3. Note that the dispersion of the regression residuals shifts to the left as θ increases in both figures.

From inspection of Table A.1, under regression quantile modeling of the data, more than 48% of the property valuations have to be raised before an appealing set of new valuations emerges. The net increase in property valuations is positive for models with $\theta \geq 0.4818$. Correspondingly, the number of potential complaints due to property valuations raised 10% or more steadily increases. At this point, tradeoffs between the two loss measures become worthy of examination and the parameter θ serves as a meaningful reference point for the possibilities. In Section 5, we discuss how to identify among the set of all regression quantile models those with attractive tradeoffs based on dominance and other considerations in the search for the final valuation model.

3. Literature review

Because the valuation of residential property may be framed as an optimization problem, a variety of modeling methods, principles (criteria), and post-optimization analyses of operations research can assist in solving the problem and analyzing the results. This is the approach taken in the paper. In addition to operations research, contributions to residential property valuation are found in the literature of econometrics, real estate finance and appraisal, statistics, and others. In this section, contributions that relate to the valuation problem as framed above are organized according to the taxonomy of Pagourtzi et al. (2003). It consists of two major categories: 1) traditional methods and 2) advanced methods. Most contributions of operations research appear in the latter.

3.1 Traditonal Valuation Methods

The Appraisal Institute publishes two popular volumes that treat residential property valuation, (Linne et al., 2000; Kane et al., 2004). Wang et al. (2002) provided a collection of essays that address valuation theory, methods, and the literature of the time.

Pagourtzi et al. (2003) included multiple regression and stepwise regression in the traditional valuation methods category of their taxonomy.

3.1.1 Least Squares, minimum sum of absolute errors, and multiple criteria regression models

When the single equation linear regression model (1) is used to valuate residential property, the ordinary least squares methodology is often used to estimate the parameters of the model, (Ihlanfeldt and Martinez-Vazquez, 1986). Several considerations account for the popularity of least squares in this context. Among others, the statistical properties of the results are well known and software is conveniently available for obtaining the least squares results. If the elements of ε in (1) are uncorrelated with expected value zero and common variance σ^2, the least squares estimator of β is best linear unbiased estimator. When the elements of ε follow the normal distribution, the least squares estimator of β is the maximum likelihood estimator.

For many practical problems, the nature of the disturbance distribution ε in (1) is rarely, if ever, known completely; the errors may not arise from a single distribution; outliers may occur but may be difficult to detect; and the choice of a loss function may not be clear from statistical, practical, or other considerations. Because properties differ in their physical characteristics and the market values of comparable single-family residences have wide variability, the disturbances in y may not arise from a single distribution. Furthermore, the consequence (loss) of implementing the property valuations derived from the model are not proportional to the squared error of prediction implicit in the least squares methodology. As noted in Section 2, when assessed property value is used to determine property taxes, the gain or loss in tax revenues due to changes in property valuations are directly proportional to the sign and the magnitude of the adjustments.

Alternatives to the least squares regression model have been proposed for modeling property values (Coleman and Larsen, 1991; Caples et al., 1997). If the elements of ε in (1) follow the Laplace distribution, the minimum sum of absolute errors (MSAE) estimator of β is the maximum likelihood estimator. Narula and Wellington (2007) proposed multiple estimation criteria for modeling residential property values.

3.1.2 Quantile regression

Unlike the least squares based regression methods, quantile regression provides a family of models that are a function of a very descriptive parameter that relates to the inherent loss associated with valuations derived from the single equation linear model (1). For this reason, it is very attractive in valuating residential property.

Koenker (2005) provided a comprehensive review of quantile regression that included statistical inference, computational methods, and other topics. Koenker and Hallock (2001) presented a compelling case for quantile regression and cited applications. Narula and Wellington (1991) framed quantile regression as a bicriteria optimization problem.

The following properties and uses of regression quantile modeling are reported in the literature:

- Unlike a unique fit provided by least squares or MSAE regression, the regression models corresponding to different values of θ provide alternate models for consideration and serve as good descriptive statistics of multifactor data (Hogg, 1975; Bassett and Koenker, 1982).
- They may be used to detect heterogeneity of the error variance (Koenker and Bassett, 1982).
- They may provide a useful way to detect outliers in a data set (Portnoy, 1982).
- They allow assignment of different weights to the positive and negative errors, which are desirable if the loss associated with over- and under-valuation is different (Reeves and Lawrence, 1982).
- Regression quantile estimators may mimic any L-estimator of location such as the median, Gastwirth's estimator, or Tukey's trimean estimator.
- Regression quantile estimators have comparable efficiency to the least squares estimators for Gaussian models while substantially outperforming the least squares estimators over a wide class of non-Gaussian error distributions. In particular, the MSAE (θ=1/2) estimator has a strictly smaller confidence ellipsoid than the least squares estimator for any disturbance distribution for which the sample median is a more efficient estimator of location than the sample mean (Koenker and Bassett, 1978).
- Regression quantiles provide a good starting solution for certain robust regression procedures. It is possible that the performance of some iterative robust regression procedures can be improved or their computational effort reduced or both by using the MSAE (θ =1/2) estimate instead of the least squares estimate as the starting solution.
- Trimmed least squares procedures that utilize regression quantiles are discussed in Koenker and Bassett (1978) and Ruppert and Carroll (1980).

3.*1.3. Other regression methods*

In addition to the above, the following regression methods are reported in the literature of residential property valuation: rank regression (Cronan et al., 1986); ridge regression (Moore et al., 2003; Ferreira and Sirmans, 1988); robust regression methods (Janssen et al., 2001). Isakson (2001) discussed the pitfalls of using (1) in real estate appraisal.

3.2 Advanced valuations methods

This category includes methodologies familiar to operations researchers such as neural networks, hedonic models, spatial analysis, fuzzy logic, and time series methods. Methodologies in the recent real estate valuation literature include: Bayesian approach (Atkinson and Crocker, 2006); goal programming (Aznar, 2007); data envelopment analysis (Lins et al., 2005); hierarchial linear model (Brown and Uyar, 2004); multiple criteria decision modeling (Fischer, 2009; Kaklauskas et al., 2007); neural networks (Peterson and Flanagan, 2009); spatial analysis and GIS (Chica-Olmo (2007); Pagourtzi et al., 2006). We also note the use of analytic hierarchy process (AHP) in house selection modeling by Ball and Srinivasan (1994) and valuation of urban industrial land using analytic network process (ANP) by Aragones-Beltran et al (2008).

4. Problem formulation

For the regression quantiles problem, let b denote the θth regression quantile estimate of β and e = y - ŷ denote the nx1 vector of corresponding regression residuals (changes in valuations) where ŷ (= Xb) is the vector of new property valuations. Consider the check function

$$\rho_\theta(\mu) = \{\theta\,\mu \text{ if } \mu \geq 0,\ (\theta - 1)\,\mu \text{ if } \mu < 0\}$$

for $\theta \in [0, 1]$. For a given value of θ, the θth regression quantile is the solution of

$$\text{Minimize } \sum_i \rho_\theta (e_i). \qquad (2)$$

where i=1,...,n. When $\theta = 0$, all residuals will be non-negative because positive residuals have zero weight in the minimization of (2). On the other hand, when $\theta = 1$, all residuals will be non-positive since they have zero weight in (2).

The θth regression quantile estimate b of β can be obtained by solving (2) iteratively. However, Koenker and Bassett (1978) have shown that b may be obtained from the solution to the equivalent (primal) linear parametric programming problem

$$\begin{array}{ll} \text{Minimize} & \theta 1' e^{+} + (1-\theta)\, 1' e^{-} \\ \text{Subject to} & Xb + e^{+} - e^{-} = y \\ & e^{+}, e^{-} \geq 0 \\ & b \text{ unrestricted in sign} \end{array} \qquad (3)$$

where $e^{+} - e^{-} = e$ and 1 is the nx1 unit vector. Note that the non-zero elements of e^{+} represent the magnitude of new valuations $\hat{y}$ (= Xb) below current values y and the non-zero elements of e^{-} indicate the contrary. Because $y - \hat{y} = e = e^{+} - e^{-}$, $e_i^{+} \cdot e_i^{-} = 0$, i = 1,...,n. The dual linear programming (LP) problem for (3) may be written as:

$$\begin{array}{ll} \text{Maximize} & y'f + (\theta-1)\, y'1 \\ \text{Subject to} & X'f = (1-\theta)\, X'1 \\ & 0 \leq f \leq 1. \end{array} \qquad (4)$$

For $\theta = ½$, the preceding formulations result in the MSAE regression model.

For any value of θ in [0,1], the solution to (3) retains essential features of the MSAE regression problem. Among those, the fitted regression quantile model passes through at least as many data points as k, the number of unknown parameters in the model (Narula and Wellington, 1986). Consequently, the number of valuations with zero adjustment ($e_i = 0$) is at least k.

From the complementary slackness property of the primal/dual relationship, $f_i = 0$ when $e_i < 0$; $0 < f_i < 1$ when $e_i = 0$; and $f_i = 1$ when $e_i > 0$, $i = 1,\ldots,n$. As θ of (3) approaches 1, the number of $f_i = 0$ in (4) increases. Consequently, for the primal problem (3), few of the regression residuals are negative near θ = 0 and few are positive near θ = 1. For the dual problem (4) and θ near zero, few responses (y) are on or below the QR hyperplane and for θ near one most responses are below the hyperplane. The duality of model estimation is clear i.e, minimizing residual error on and about the QR hyperplane is equivalent to sectioning the X, y space so that as many responses as possible lie on/below the desired QR hyperplane. The duality is subtle but helpful in understanding how QR modeling and its parameterization under θ position the regression hyperplane as a quantile estimate.

When the intercept term is included in the model, the dual problem (4) includes the constraint

$$\sum_i f_i = (1-\theta)\, n \qquad (5)$$

that is, the average reduced cost is (1- θ) and

$$\sum_i f_i = n^+ + \sum_p f_p \qquad (6)$$

$$\theta = [\ (k - \sum_p f_p) + n^- \ ] / n \qquad (7)$$

where n^+ is the number of positive residuals and the f_p are the dual variables for which $p = \{i \mid 0 < f_i < 1, i = 1, \ldots, n\}$.

Interestingly,

$$\hat{y}'f = y'f + e'f. \qquad (8)$$

For a specified value of θ, the primal linear programming problem (3) can be solved efficiently using a slightly modified version of the Barrodale and Roberts (1973) algorithm for solving the MSAE regression problem. Computer programs given in Wellington and Narula (1984) and

Koenker and D'Orey (1987) may be used to compute all regression quantiles associated with a data set. Koenker (2006) provided an R language implementation called quantreg for various quantile regression methods. SPSS version 17 and above as well as the STATA software packages include routines for quantile regression. They do not provide a convenient option for suppressing the intercept term in the model.

5. Model selection and analysis

The preferred θth regression quantile model(s) should produce changes in property valuations that cause few challenges by property owners and should offer some gain in aggregate property valuations. For those regression quantile models, the net gain in the tax base, $1'e^- - 1'e^+$, will be as large as possible with as few as possible potential challenges from property owners. Challenges are likely for valuations raised 10% or greater and more likely for those raised 20% or more. In Table A.1 of the Appendix, we recorded the net increase in property valuations and the number of valuations increased at or above thresholds of 10% and 20% for each regression quantile model of the real estate data discussed in Section 2. Among the entries, note for some regression quantile models the net increase in valuations is smaller or the number of valuations increased at or above thresholds of 10% and 20% are greater than other models, i.e., they are dominated. The non-dominated regression quantiles are presented in Table 2 and provide a reduction in the number of candidates for the final model selection.

The decision maker may have a goal to increase valuations among the fifty-four properties by at least \$0.25M, \$0.50M, \$0.75M, or \$1.0M and in so doing provoke as few challenges to new property valuations as possible. For these situations, QR models for θ = 0.5078, 0.6667, 0.7637, and 0.8471 respectively may have appeal, see Table 2. Among these models, the consequences (net change in property valuations and number of potential challenges) of having more (or fewer) increased property valuations are readily determined from the inspection of adjacent regression quantile model(s). For example, the θ = 0.6667 regression quantile model for the real estate data

$$\hat{y} = 58.247x_1 + 8.804x_2 - 0.099x_3 + 2.354x_4 - 13.853x_5 - 7.932x_6 + 2.133x_7 - 0.784x_8 + 0.213x_9 + 7.203x_{10} \quad (9)$$

will increase nearly 67% of the current property valuations. For this model, the number of new valuations at or above thresholds of 10% and 20% is respectively 18 and 9. For an additional valuation above 20%, the adjacent model for θ = 0.6846 offers an additional $76,247 in net property valuations. The fitted regression quantile model for θ = 0.6846 is

$$\hat{y} = 57.342x_1 + 7.787x_2 - 0.182x_3 + 2.944x_4 - 18.915x_5 - 5.819x_6 + 7.107x_7 - 7.229x_8 + 0.207x_9 + 5.449x_{10}. \quad (10)$$

In another comparison, the model for θ = 0.7472 offers $105,638 increase in net property valuations with one additional valuation above 10% and one above 20% relative to the model for θ = 0.6667. Comparisons such as these among the non-dominated QR models assist the decision maker in identifying an appealing final model.

In finalizing the valuation model, it is important for the analyst to understand which residential properties form the basis for the valuations resulting from the final model. In this regard, consider the following. Each binding constraint of the optimal solution to (3) under any θ in [0, 1] corresponds to a regression residual ($e_i^+ - e_i^-$) with value zero and accordingly an unchanged property valuation. For each QR model, there are at least k (= number of elements of b) such constraints or residential properties. As a consequence, the elements of b for each QR model are completely determined by the corresponding binding constraints, that is,

$$b = X_1^{-1} y_1 \quad (11)$$

where X_1 is the kxk array containing the X-data of the binding constraints and y_1 is the kx1 vector of corresponding values in y. Clearly, the binding constraints and the residential properties (observations) they relate to are most influential. They may be looked upon as a reference set. Challenges to valuations may be explained by the difference in features between those of the reference set and the contested property. It is also useful to identify which properties are influential (binding) at the extremes and the center of the data.

Figure 4 is a display of the instances in which each constraint is binding among the various regression quantile models of the real estate data set. Let i(•) denote the number of regression quantile models in which constraint i is binding, i=1,...,54. Constraint 46(85) is binding most often and constraints 7(1) and 29(1) are the least binding among the one hundred fifteen regression quantile models. For models with $\theta \geq 0.5$, constraint 46 is binding. Constraint 32 is binding for all models with $0.1316 \leq \theta \leq 0.5047$ and for $0.1942 \leq \theta \leq 0.6667$ it is binding with either constraints 30 or 31 or both. Because observation 46 is among the k = 10 binding constraints that determine $b_1, \ldots, b_{10}$ in eighty-five of the one hundred fifteen regression quantile models of the real estate data set, it may be an outlier, Portnoy (1982) and Bassett and Koenker (1982). The same may be true for constraints 5(63), 30(61), and 32(67). Due to the role these observations play in determining the vectors of parameter estimates, the analyst may want to confirm each datum of the observations.

Observe in Figure 4 that the sets of binding constraints for any two adjacent quantile models differ by one element. It is possible to generate all QR models for the real estate data beginning with the solution for an initial QR model (θ = 0 or 1) and selectively exchanging one element in its set of binding constraints with an element from its non-binding constraint set. Each exchange of this kind and the solution (QR model) it generates may be achieved with a simplex iteration. Successive exchanges produce all QR models.

6. Remarks

Unlike the least squares regression methodology, quantile regression provides a family of models that are a function of a very descriptive parameter that relates to the inherent loss associated with residential property valuations derived from the single equation multiple linear regression model. In property valuation of this kind, we argued that loss/gain in tax base and increase/decrease in challenges of property owners arising from changes in valuations are proportional to the sign and magnitude (both absolute and relative) of the regression residual. We found quantile regression modeling to be well suited to addressing loss of this kind. We showed how the set of all possible regression quantile models for a data set can be reduced to a smaller set of attractive models using the principle of dominance; how to examine tradeoffs among loss measures associated with appealing regression quantile models; how to analyze the LP solutions to the QR problem; and suggested displays such as Table 2 and Figures 1 - 4 to assist in final model selection. We illustrated the approach with a data set.

The approach incorporates methodology of single and multiple criteria optimization in generating and analyzing meaningful alternate models for valuating residential property and as such is good illustration of the interfacing of operations research and data analysis.

The approach may be adapted to the valuation of other assets such as robot technologies and the rating of vendors. In each case, the valuation may be based on the asset's characteristics (X) and its corresponding value (y). The elements of y could be the asset cost or a measure of its effectiveness. The valuation derived from the regression approach can over-state or under-state an asset's value and correspondingly enhance or lessen correctly or otherwise the appeal of the assets so valuated. Suppose an analyst is confronted with evaluation of competing robot technologies with varying characteristics such as maximum load or lifting capacity, velocity, repeatability, acquisition cost as well as computer related features such as memory, processing speed, image display, and others, (Imang and Schlesinger, 1989; Rao and Padmanabhan, 2006; Chatterjee et al., 2010). Let repeatability (y, precision with which a robot returns to a given point under a specified load and velocity) be prime concern and be related to the other technical features (X) in a hedonic linear regression model for a set of competing robots. The positive (negative) regression residual indicates under- (over-) achievement of repeatability, i. e. loss. The former is poor performance and the latter is not possible. Robots that are consistently fitted with zero residual error at or about the middle/central (e.g. $0.4 < \theta < 0.6$) regression quantiles standout in the modeling and may constitute a first subset of robots to be evaluated for pre-purchase investigation and experimentation. If the set of competing robot technologies is large (Rao and Padmanabhan, 2006), pre-purchase testing is expensive, and if the hedonic model is correct, QR modeling of the data provides means for identifying a meaningful reduction in the number technologies for evaluation.

Acknowledgements

We thank Trudy Reiser for providing us with the data for the example. We also thank the staff of the County Assessor Office of Allen County Indiana for many insightful observations relating to residential property valuation. The work of Professor Narula was partially supported by the Summer Research Grant, School of Business, Virginia Commonwealth University, Richmond, Virginia. We thank the referees for many suggestions and insights that improved the presentation.

References

Aragones-Beltran, P., Aznar, J., Ferris-Onate, J., Garcia-Melon, M., 2008. Valuation of urban industrial land: an analytic network process approach. European Journal of Operational Research, 185 (1), 322-329.

Atkinson, S., Crocker, T., 2006. A Bayesian approach to assessing the robustness of hedonic property value studies. Journal of Applied Econometrics 2 (1), 27-45.

Aznar, J., Guijarro, F., 2007. Estimating regression parameters with imprecise input data in an appraisal content. European Journal of Operational Research, 176 (3), 1896-1907.

Ball, J., Srinivasan, V., 1994. Using the analytic hierarchy process in house selection. Journal of Real Estate Finance and Economics 9, 69-85.

Barrodale I., Roberts, F.D.K., 1973. An improved algorithm for discrete L_1 linear approximation. SIAM Journal on Numerical Analysis 10, 839-848.

Bassett G., Koenker R., 1982. An empirical quantile function for linear models with iid errors. Journal of the American Statistical Association 77, 407-415.

Brown, K., Uyar, B., 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education 7 (1), 15-23.

Caples S., Hanna, M., Premeaux, S., 1997. Least squares versus least absolute value in real estate appraisal. The Appraisal Journal (January), 17-24.

Chatterjee, P., Athawale, V., Chakraborty, S. 2010. Selection of industrial robots using compromise ranking and outranking methods. Robotics and Computer-Integrated Manufacturing, 26, 483-489.

Chica-Olmo, J., 2007. Prediction of housing location price by a multivariate spatial method: Cokriging. Journal of Real Estate Research 29 (1), 91-114.

Coleman, J., Larsen, J., 1991. Alternative estimation techniques for linear appraisal models. The Appraisal Journal (October), 526-532.

Cronan, T., Epley, D., Perry, L., 1986. The use of rank transformation and multiple regression analyss in estimating residential property values with a small sample. The Journal of Real Estate Research 1(1), 19-31.

Ferreira, E., Sirmans, G., 1988. Ridge regression in real estate analysis. The Appraisal Journal, 56 (3), 311- 319.

Fischer, D., 2009. Multiple criteria decisions: opening the black box. Journal of Real Estate Practice and Education 12 (1), 43-55.

Hogg, R.V., 1975. Estimates of percentile regression lines using salary data. Journal of the American Statistical Association 70, 56-59.

Ihlanfeldt, K., Martinez-Vazquez, J., 1986. Alternative value estimates of owner-occupied housing: evidence on sample selection bias and systematic errors. Journal of Urban Economics 20, 356-369.

Imang, M., Schlesinger, R. 1989. Decision models for robot selection: a comparison of ordinary least squares and linear goal programming methods. Decision Sciences, 20, 40 - 53.

Isakson, H.R., 2001. Using multiple regression in real estate appraisal. The Appraisal Journal (October), 424-430.

Janssen, C., Soderberg, B., Zhou, J., 2001. Robust estimation of hedonic models of price and income for investment property. Journal of Property Investment & Finance 19(4), 342-360.

Kaklauskas, A., Zavadskas, E., Banaitis, A., Satkauskas,G., 2007. Defining the utility and market value of a real estate: a multiple criteria approach. International Journal of Strategic Property Management 11 (June), 107-120.

Kane, M.S., Linne, M., Johnson, A., 2004. Practical applications in appraisal valuation modeling. Appraisal Institute, Chicago.

Koenker, R.W., 2005. Quantile regression. Cambridge University Press, New York.

Koeneker, R.W., 2006. Quantile regression in R: a vignette. Available at http://www.econ.uiuc.edu/~roger/research/rq/vig.pdf.

Koenker, R.W., Bassett, G.W., 1978. Regression quantiles. Econometrica 46 (1), 33-50.

Koenker, R.W., Bassett, G.W., 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50 (1), 43-61.

Koenker, R.W., d'Orey, V., 1987. Computing regression quantiles. Applied Statistics 36, 383-393.

Koenker, R.W., Hallock K., 2001. Quantile regression. Journal of Economic Perspectives 15 (4), 143–156.

Kummerow, M., 2000. Thinking statistically about valuations. The Appraisal Journal. 68 (3), 318-326.

Linne, M., Kane, M.S., Dell, G., 2000. A guide to appraisal valuation modeling. Appraisal Institute, Chicago.

Lins, M., Novaes, L., Legey, L., 2005. Real estate appraisal: a double perspective data envelopment analysis approach. Annals of Operations Research 138 (1), 79-96.

Moore, J., Reicheri, A., Cho, C., 2003. Analyzing the temporal stability of appraisal model coefficients: an application of ridge regression techniques. Real Estate Economics 12 (1), 50-71.

Narula, S., Wellington, J., 1986. An algorithm to find all regression quantiles. Journal of Statistical Computation and Simulation 26 (3-4), 269-281.

Narula, S., Wellington, J., 1991. An algorithm to find all regression quantiles using bicriteria optimization. American Journal of Mathematical and Management Sciences 10, 229-259.

Narula, S., Wellington, J., 2007. Multiple criteria linear regression. European Journal of Operational Research 181 (2), 767-772.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, F. N., 2003. Real estate appraisal: a review of valuation methods. Journal of Property Investment & Finance 21 (4), 383-401.

Pagourtzi, E., Nikolopoulos, K., Assimakopoulos, V., 2006. Architecture for a real estate analysis information system using GIS techniques integrated with fuzzy theory. Journal of Property Investment & Finance 24 (1), 68-78.

Peterson, S., Flanagan, A., 2009. Neural network hedonic pricing models in mass real estate appraisal. Journal of Real Estate Research 31(2), 147-164.

Portnoy, S., 1982. Book review of Robust Statistics by P. J. Huber. Technometrics 24, 163-164.

Rao, R. V., Padmanabhan, K. 2006. Selection, identification and comparison of industrial robots using digraph and matrix methods. Robotics and Computer-Integrated Manufacturing, 22, 373-383.

Reeves, G.R., Lawrence, K.D., 1982. Combining multiple forecasts given multiple objectives. Journal of Forecasting 1, 271-279.

Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 75, 828-838.

Wang, K., Wolverton, M. (eds.), 2002. Real estate valuation theory. Kluwer Academic Publishers, Norwell, MA.

Wellington, J., Narula, S., 1984. An algorithm for regression quantiles. Communications in Statistics 13B, 683-704.

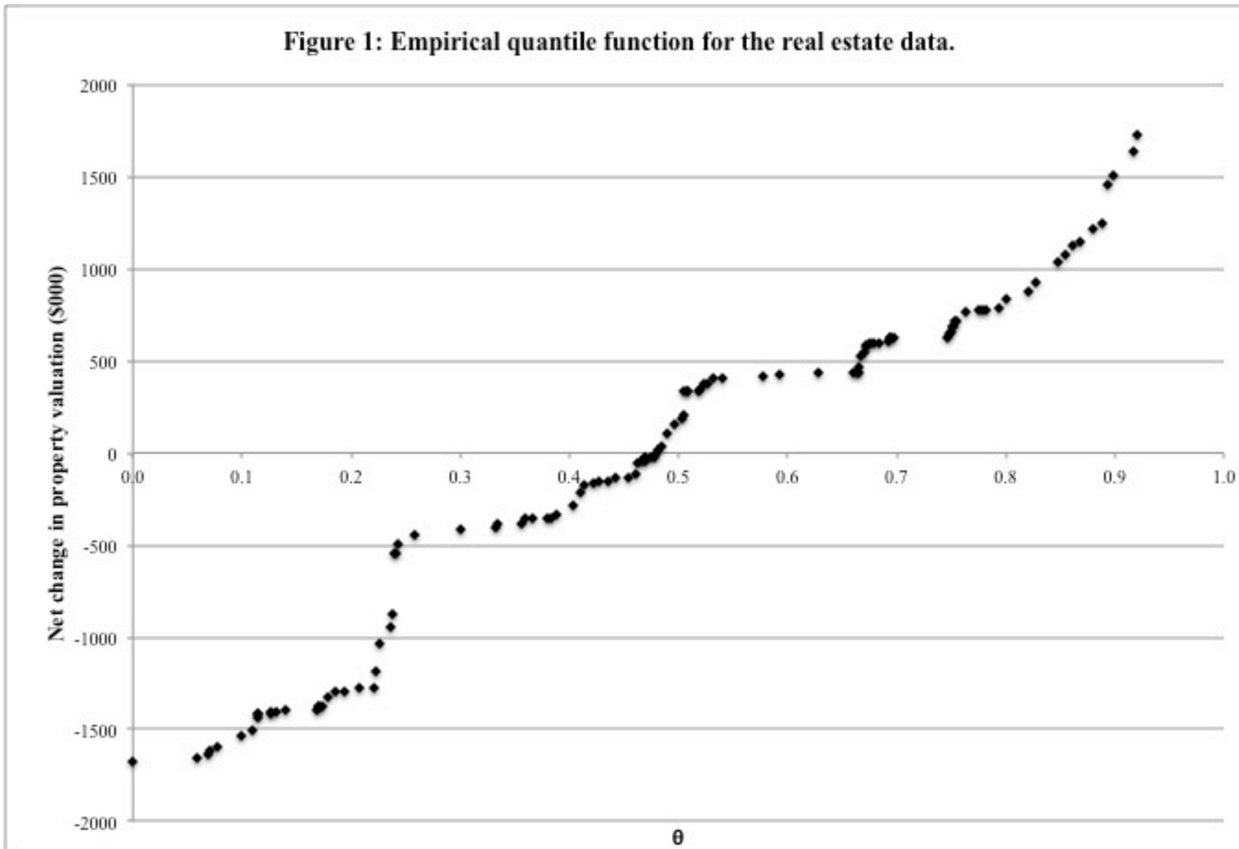
Figure 1: Empirical quantile function for the real estate data.
Net change in property valuation ($000)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
θ

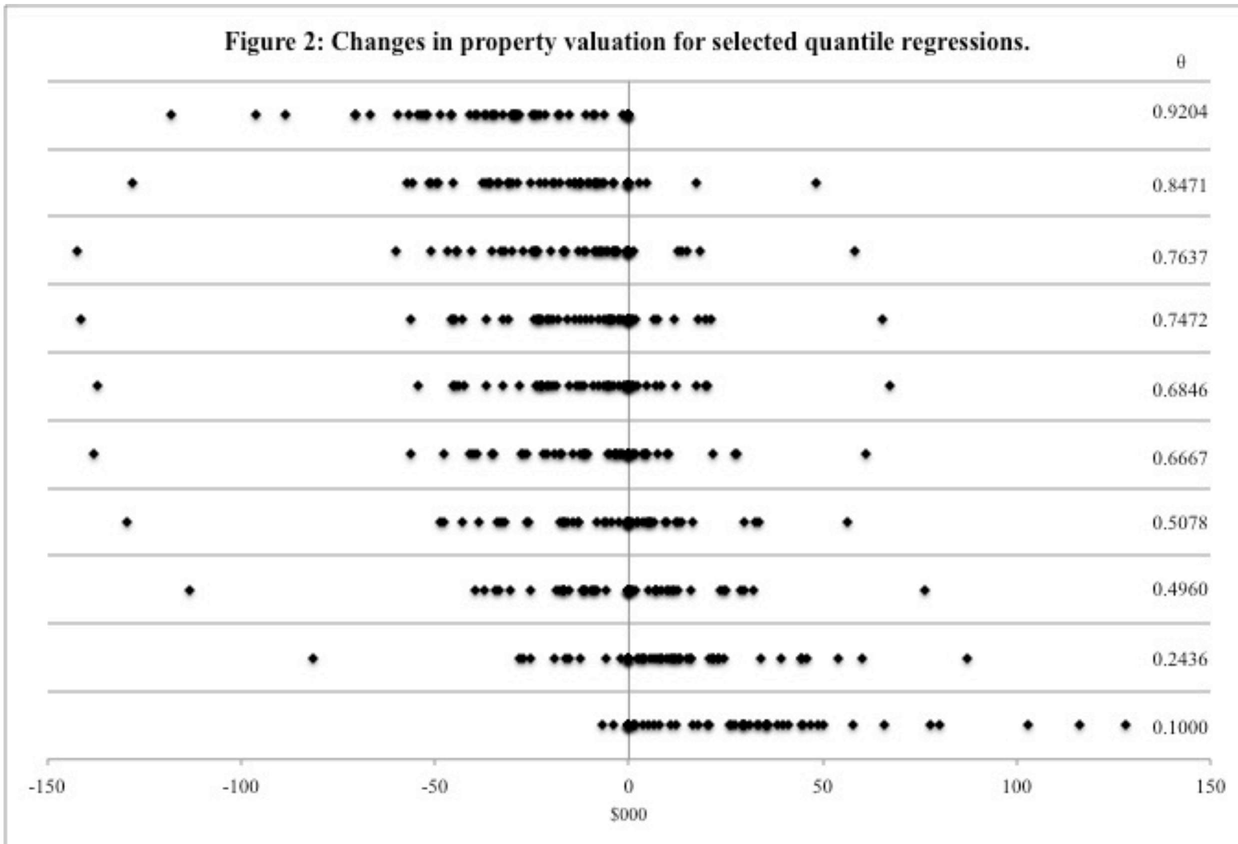
Figure 2: Changes in property valuation for selected quantile regressions.
θ
0.9204
0.8471
0.7637
0.7472
0.6846
0.6667
0.5078
0.4960
0.2436
0.1000
-150
-100
-50
0
50
100
150
$000

Figure 3: Percentage changes in property valuations for selected quantile regressions.

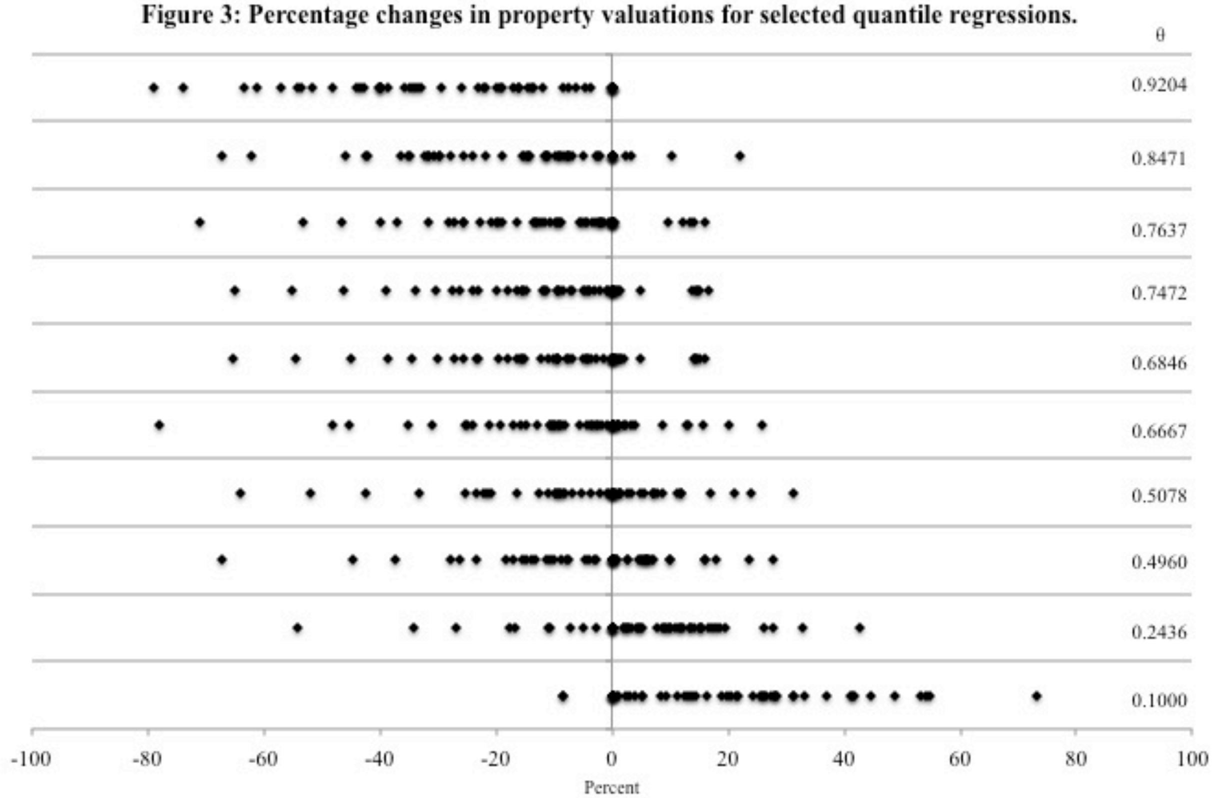

Figure 4: Observations with zero residual error for all quantile regression models.

	Observation																																																				
θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
0.9204					5		7			10								18			21												33													46			49			52	53	
0.9170					5					10								18			21												33													46			49			52	53	54
0.8986					5	6				10								18			21												33													46						52	53	54
0.8929			3		5	6												18			21												33													46						52	53	54
0.8885			3		5	6												18			21												33													46					51	52		54
0.8798			3		5	6												18			21												33													46			49		51			54
0.8674			3		5	6														20	21												33													46			49		51			54
0.8606			3		5	6												18		20	21												33													46			49		51			
0.8540			3		5	6												18		20													33				37									46			49		51			
0.8471			3		5	6														20													33				37									46			49		51			54
0.8276			3		5	6															21												33				37									46			49		51			54
0.8203			3		5	6															21				25								33													46			49		51			54
0.8007			3		5	6															21				25								33				37									46					51			54
0.7929			3		5	6															21				25								33				37			40						46								54
0.7812			3		5																21				25							32	33				37			40						46								54
0.7786			3						9												21				25							32	33				37			40						46								54
0.7745					5				9												21				25							32	33				37			40						46								54
0.7637					5	6			9												21				25							32	33							40						46								54
0.7549					5	6			9										19						25							32	33							40						46								54
0.7526			3		5	6			9										19						25							32	33													46								54
0.7518			3		5				9			12							19						25							32	33													46								54
0.7507					5				9			12							19						25						31	32	33													46								54
0.7490					5				9			12							19				23		25						31		33													46								54
0.7480					5				9										19				23		25						31		33					38								46								54
0.7472					5				9			12							19				23		25						31							38								46								54
0.6978					5				9			12							19				23								31						37	38								46								54
0.6947					5	6			9			12							19				23								31							38								46								54
0.6942					5	6			9			12							19				23							30	31							38								46								
0.6924					5				9			12							19				23							30	31							38								46			49					
0.6846					5				9			12							19				23		25						31							38								46			49					
0.6782					5				9			12							19				23		25													38								46			49			52		
0.6755					5	6			9			12							19				23															38								46			49			52		
0.6726			3		5	6			9										19				23															38								46			49			52		
0.6722			3		5	6			9			12											23															38								46			49			52		
0.6710			3		5	6			9			12											23									32						38								46			49					
0.6667			3		5	6			9														23							30		32						38								46			49					
0.6655			3	4	5	6			9														23							30		32														46			49					
0.6650			3	4	5				9														23							30		32														46			49			52		
0.6642			3	4	5																		23							30	31	32														46			49			52		
0.6604			3	4	5																		23				27			30	31	32														46			49					
0.6288			3	4	5																		23				27				31	32														46			49			52		
0.5929			3	4	5																		23				27				31				35											46			49			52		
0.5772			3	4	5																		23					28			31				35											46			49			52		
0.5402				4	5																		23					28		30	31				35											46			49			52		
0.5318				4																			23					28		30	31				35							42				46			49			52		

Figure 4: Observations with zero residual error for all quantile regression models. (Continued.)

Observation

θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
0.5262			3	4																			23					28		30	31				35							42				46						52		
0.5235			3	4					9																			28		30	31				35							42				46						52		
0.5200			3	4					9								17													30	31				35							42				46						52		
0.5178			3	4													17													30	31				35			38				42				46						52		
0.5081			3														17						23							30	31				35			38				42				46						52		
0.5078				4													17						23							30	31				35			38				42				46						52		
0.5047				4													17						23							30	31	32			35							42				46						52		
0.5044				4	5												17						23							30	31	32			35							42				46								
0.5039					5				9								17						23							30	31	32			35							42				46								
0.4960					5				9								17						23					28		30	31	32			35							42												
0.4896					5				9			12					17						23					28		30	31	32										42												
0.4838			3		5				9			12					17						23							30	31	32										42												
0.4818			3		5				9			12					17						23							30	31	32							39															
0.4802			3		5				9			12					17						23							30		32							39			42												
0.4778					5				9			12					17						23							30		32							39			42			45									
0.4747					5				9			12											23							30		32							39		41	42			45									
0.4701	1				5				9			12											23							30		32							39		41	42												
0.4700	1		3		5							12											23							30		32							39		41	42												
0.4699	1		3		5							12											23							30	31	32							39		41													
0.4662	1				5				9			12											23							30	31	32							39		41													
0.4633	1		3		5				9			12																		30	31	32							39		41													
0.4613	1		3		5				9			12																		30	31	32									41					46								
0.4533			3		5				9			12																		30	31	32									41					46		48						
0.4415			3		5							12			15															30	31	32									41					46		48						
0.4360			3		5										15	16														30	31	32									41					46		48						
0.4269	1		3		5										15	16														30	31	32														46		48						
0.4227	1				5										15	16														30	31	32									41					46		48						
0.4135	1								9						15	16														30	31	32									41					46		48						
0.4104	1								9						15	16														30	31	32									41				45			48						
0.4038	1			4					9						15															30	31	32									41				45			48						
0.3889		2		4					9						15															30	31	32									41				45			48						
0.3827		2							9						15															30	31	32							39		41				45			48						
0.3792		2							9						15	16															31	32							39		41				45			48						
0.3664		2							9		11				15	16															31	32							39		41				45									
0.3589		2							9		11				15															30	31	32							39		41				45									
0.3562		2		4					9		11				15															30	31	32							39						45									
0.3347		2		4					9		11				15																31	32							39						45					50				
0.3323		2		4					9		11				15															30	31	32							39											50				
0.3011		2							9		11				15	16														30	31	32							39											50				
0.2587		2									11				15	16						22								30	31	32							39											50				
0.2436		2									11			14	15	16						22								30		32							39											50				
0.2410		2									11			14	15	16						22								30		32														46				50				
0.2403		2							9		11			14	15	16						22								30		32														46								
0.2382		2							9		11			14	15	16						22								30		32											43											
0.2365		2							9		11			14		16						22								30		32											43			46								

Figure 4: Observations with zero residual error for all quantile regression models. (Continued.)

Observation

θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
0.2254		2									11			14		16						22								30		32											43	44		46								
0.2228											11			14		16						22								30		32				36							43	44		46								
0.2204											11			14		16						22								30		32		34		36								44		46								
0.2083								8			11			14		16						22								30		32		34										44		46								
0.1942								8			11			14		16														30		32		34	35									44		46								
0.1860								8			11			14		16																32		34	35								43	44		46								
0.1792								8			11			14		16																32		34	35	36							43			46								
0.1734								8			11					16						22										32		34	35	36							43			46								
0.1707								8								16						22				26						32		34	35	36							43			46								
0.1697								8			11					16						22				26						32		34	35								43			46								
0.1678											11		13			16						22				26						32		34	35								43			46								
0.1397													13			16						22				26						32		34	35								43			46	47							
0.1316								8					13			16						22				26						32			35								43			46	47							
0.1269								8			11		13			16						22				26									35								43			46	47							
0.1264								8			11		13			16										26				30					35								43			46	47							
0.1147								8			11		13			16														30					35	36							43			46	47							
0.1144								8			11		13			16						22													35	36							43			46	47							
0.1142								8					13			16						22												34	35	36							43			46	47							
0.1100								8					13			16																		34	35	36							43	44		46	47							
0.1000								8					13			16														30				34		36							43	44		46	47							
0.0771								8					13			16								24						30				34									43	44		46	47							
0.0708								8					13			16								24						30				34		36							43	44			47							
0.0685				4				8					13			16								24										34		36							43	44			47							
0.0592				4				8					13			16								24										34	35								43	44			47							
0.0000				4				8					13			16								24					29					34	35									44			47							

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

Table 2: The loss measures associated with the non-dominated regression quantile models.

	θ	Max % Change Among Valuations	Net Gain In Valuations ($000)	No. of Valuations Raised More Than 10%	20%
1	0.9204	106.75	1722.671	38	29
2	0.9170	102.66	1638.563	37	27
3	0.8986	96.39	1510.047	34	25
4	0.8929	93.86	1455.640	33	25
5	0.8885	81.37	1243.673	33	23
6	0.8798	79.36	1218.182	32	23
7	0.8674	74.42	1143.122	31	20
9	0.8540	66.03	1077.259	31	19
10	0.8471	67.26	1033.722	28	18
11	0.8276	67.22	922.043	24	18
12	0.8203	69.74	872.428	23	17
13	0.8007	70.38	832.800	23	16
14	0.7929	70.54	784.260	21	14
18	0.7637	71.02	764.018	22	12
19	0.7549	69.83	719.159	21	11
23	0.7490	65.35	650.609	20	11
25	0.7472	65.07	629.592	19	10
30	0.6846	65.32	600.201	18	10
36	0.6667	78.09	523.954	18	9
37	0.6655	80.39	461.039	15	9
38	0.6650	68.86	439.130	15	8
46	0.5262	64.81	379.277	14	8
48	0.5200	63.65	343.581	13	9
51	0.5078	63.85	334.786	12	9
53	0.5044	67.84	207.551	13	6
57	0.4838	67.07	38.541	11	4

[1] Obtained from Table A.1 of the Appendix.