Quantile Regression Estimates of Hong Kong Real Estate Prices

Stephen W.K. Mak, Lennon, H.T. Choy and Winky K.O. Ho
Department of Building and Real Estate
Hong Kong Polytechnic University

May 2010
Hong Kong

©2010 by Stephen Mak, Lennon Choy and Winky Ho. All rights reserved.

This paper cannot be quoted without explicit permission provided that full credit, including © notice, is given to the source.
Quantile regression estimates of Hong Kong real estate prices

Abstract. Linear regression is a statistical tool used to model the relation between a set of housing characteristics and real estate prices. It estimates the mean value of the response variable, given levels of the predictor variables. The quantile regression approach complements the least squares by identifying how differently real estate prices response to a change in one unit of housing characteristic at different quantiles, rather than estimating the constant regression coefficient representing the change in the response variable produced by a one-unit change in the predictor variable associated with that coefficient. It estimates the implicit price for each characteristic across the distribution of prices, and allows buyers of higher-priced properties to behave differently from buyers of lower-priced properties even if they are within one single housing estate. Thus, it better explains the real world phenomenon, and offers a more comprehensive picture of the relationship between housing characteristics and prices.

1. Introduction

Residential property is a multi-dimensional commodity which can be considered as a bundle of utility-bearing attributes that consumers value. These attributes are characterized by their physical inflexibility, durability, and spatial fixity such that different combinations of them can produce a heterogeneous good. In the real estate literature, housing price is defined as a function of a bundle of inherent attributes (i.e., flat size, age, floor, and balcony), neighborhood characteristics (i.e., view), accessibility (i.e., transport, the presence of recreation facilities, community services, and school), and environmental quality (waterfront or natural beauty) that yield utility or satisfaction to homebuyers. Particularly, a hedonic price model involves first the specification of a housing price function which relates the observed housing expenditure to the selected physical, neighborhood, and accessibility characteristics that are considered to influence prices (Bailey et al., 1963; Ridker and Henning, 1967; Kain and Quigley, 1970; Wilkinson, 1973; Freeman, 1979; Pollakowski, 1982; Epple, 1987; Case, 1992; Chesire and Sheppard, 1995; Can and Megbolugbe, 1997). Based
on the estimated coefficients of housing attributes, the second stage is to construct price indexes.

Linear regression is a statistical tool used to model the relation between a set of predictor variables and a response variable. It estimates the mean value of the response variable for given levels of the predictor variables. Suppose we are interested in investigating the relationship between housing prices and a set of predictors, such as apartment size, age, floor level, view, direction and car park. The data set used for this example contains a total of 5,947 cross sectional inter-temporal transaction data from City One Sha Tin, a private housing estate located in Sha Tin, the New Territories, for the January 1997 – October 2004 period. The linear regression model for this example is as follows:

\[ P = -819906.8 + 6980.7GFA - 2.8GFA^2 - 62374.2AGE + 2328.3AGE^2 + 16422.8FL - 336.8FL^2 - 21157.5BUILDING - 16767.3OBSTRUCTIVE + 73641.3NE + 7654.7SE + 82157.7SW + 166604.6CP \]

\[
\begin{align*}
&({-14.34}) \quad ({41.11}) \quad ({-18.73}) \quad ({-11.69}) \quad (13.87) \\
&({16.36}) \quad ({-11.49}) \quad (-2.46) \\
&({-1.78}) \quad (12.57) \quad (1.48) \\
&({15.35}) \quad (31.09) \\
\end{align*}
\]

This model estimates how, on average, these properties’ characteristics impact on real estate prices. The car park predictor variable, \( CP \), compares the effect of having a car park on property prices with not having a car park. While this model can address the question of whether or not a car park matters in the price determination, it cannot answer another important question: “Does a car park influence property prices differently for low-priced properties than for median-priced properties?” One can obtain a more comprehensive picture of the effect of the predictors on the response

\[ \text{For the data definitions and sources, please refer to Section 4 Data Sources.} \]
variable by using quantile regression, which models the relation between a set of predictor variables and the specific percentiles (or quantiles) of the response variable. It specifies changes in the quantiles of the response. For example, a median regression (the 50th percentile) of property prices on properties’ characteristics specifies the changes in the median property prices as a function of the predictors. The effect of car park on median property prices can be compared to its effect on other quantiles of property prices.

In linear regression, the regression coefficient represents the change in the response variable produced by a one-unit change in the predictor variable associated with that coefficient. The quantile regression parameter estimates the change in a specified quantile of the response variable produced by a one-unit change in the predictor variable. This allows for a comparison of how specific percentiles of property prices may be more affected by certain properties’ characteristics than other percentiles. This is reflected in the change in the size of the regression coefficient.

The objective of this paper is to empirically estimate how specific quantiles of property prices respond differently to a one-unit change in the properties’ characteristics. As an alternative to OLS regression, this study adopts quantile regression to identify the implicit prices of housing characteristics for the different percentiles of the distribution of housing prices. This explicitly allows higher-priced apartments to have different implicit prices to a property’s characteristic than lower-priced apartments. Heckman (1979) suggests that the issues associated with truncation could possibly be avoided since quantile regression makes use of the entire sample rather than the mean value of the response variable. This will eliminate the problem of biased estimates that is created when OLS is applied to housing price subsamples (Newsome and Zietz, 1992).
This paper is organized as follows. Section 2 briefly presents a literature review of the quantile regression. Section 3 discusses the model specification adopted in this paper, while the data quality and sources will be presented in Section 4. Section 5 presents and discusses the empirical results, utilizing housing transaction data from one mega-scale housing estate, the City One Sha Tin, located in Shatin, the New Territories for the period between January 1997 and October 2004. The last section summarizes the major findings.

2. Literature Review

Quantile regression is based on the minimization of weighted absolute deviations for estimating conditional quantile (percentile) functions (Koenker and Bassett 1978; Koenker and Hallock 2001). For the median (quantile = 0.5), symmetric weights are used, while asymmetric weights are employed for all other quantiles (e.g., 0.1, 0.2, ..., 0.9). While the classical OLS regression estimates conditional mean functions, quantile regression can be employed to explain the determinants of the dependent variable at any point of the distribution of the dependent variable. For hedonic price functions, quantile regression makes it possible to statistically examine the extent to which housing characteristics are valued differently across the distribution of housing prices. Although one may argue that the same goal may be accomplished by utilizing the price series subsamples according to its unconditional distribution and then applying OLS to the subsamples, Heckman (1979) argues that the “truncation of the dependent variable” may create biased parameter estimates and should be avoided if possible. Since quantile regression employs the full data set, a sample selection problem does not arise in the first place.

Koenker and Hallock (2001) suggest that there is a rapidly expanding empirical
quantile regression literature in economics that, when taken as a whole, makes a persuasive case for the value of “going beyond models for the conditional mean” in empirical economics. This methodology has been intensively applied to the issues in labor economics, such as union wage effect, returns to education, and labor market discrimination. Chamberlain (1994) finds that for manufacturing workers, the union wage premium at the first decile is 28 percent, and declines monotonically to a 0.3 percent at the upper decile. The least squares estimate of the mean union premium of 15.8 percent is thus captured mainly by the lower tail of the conditional distribution. Other studies exploring these issues in the labor market include the influential work of Buchinsky (1994; 1997), Schultz and Mwabu (1998), and Kahn (1998). Particularly, the work of Machado and Mata (1999) is notable, since it introduces a useful way to extend the counterfactual wage decomposition approach of Oaxaca (1973) to quantile regression, and provides a general strategy for simulating marginal distributions from the quantile regression process. Arias, et al (2001), employing data on identical twins, interpret observed heterogeneity in the estimated returns to education over quantiles as an indicator of an interaction between observed educational attainment and unobserved ability.

In demand analysis, Deaton (1997) offers an introduction to quantile regression. Employing food expenditure data from Pakistan, his study finds that although the median Engel elasticity of 0.906 is similar to the ordinary least squares estimate of 0.909, the coefficient at the tenth quantile is 0.879 and the estimate at the 90th percentile is 0.946, indicating a pattern of heteroskedasticity. In another demand application, Manning, et al (1995) investigate the demand for alcohol using survey data from the National Health Interview Study, and suggest the presence of
considerable heterogeneity in the price and income elasticities over the full range of the conditional distribution.

Utilizing the American Housing Survey data, Gyourko and Tracy (1999) adopt the quantile regression approach to investigate changes in housing affordability between 1974 and 1997. Without controlling for changes over time in housing characteristics, real house prices in 1997 had risen by 35% at 0.9 quantile and had fallen by 28% at 0.1 quantile, while median prices did not change. Controlling for changes in housing characteristics over time, real house price with 1974 characteristics increased by only 1% at 0.9 quantile, while real prices increased by 33 percent at 0.1 quantile. The quantile estimates indicate that real house prices with 1974 characteristics at 0.9 quantile increased by about 31% over the time period, which is much closer to the price increase for those situations when changes over time in housing characteristics are not controlled. Real housing prices with 1974 characteristics increased by about 20% at the 0.1 quantile, less than that indicated from the mean-based estimates, but much more than for those situations when changes over time in housing characteristics are not controlled. These results suggest that quantile effects are important, while “average quality has worsened at the bottom of the house price distribution.”

Employing housing transaction data from Chicago in 1993 through 2005, McMillen and Thorsnes (2006) suggest that quantile regression has advantages over the conventional mean-based approaches to estimate housing price index. A median-based quantile estimator which reduces the outlier effect, suffers less bias from unobserved renovations than a standard mean-based estimator. The problem of outliers is particularly important for the repeat-sales estimator, which is vulnerable to an upward bias when the sample includes renovated houses and there is no way to
identify which homes have been upgraded. In this situation, a more realistic view of the housing market may be gained by constructing indexes using lower quantiles as the target point. Zietz et al (2008) utilize quantile regression, with and without accounting for spatial autocorrelation, to identify the coefficients of a large set of diverse variables across different quantiles. Their results suggest that while buyers of higher-priced homes value square footage and the number of bathrooms differently from buyers of lower-priced homes, their study finds that other variables, such as age, also vary across the distribution of housing prices.

To the best of our knowledge, our paper is the first of its kind to use quantile regression technique, based on Hong Kong housing transaction data, to investigate the implicit prices of housing characteristics in different quantiles of prices. Hong Kong, for many reasons, presents an interesting case. It is a densely populated territory, with the majority of its citizens residing in housing estates instead of standalone residential buildings or houses. Frequent transactions of residential properties within even one single housing estate (typically with 20-30 blocks of buildings) over time provide researchers with adequate observations (from a sample of similar location-specific characteristics) to employ quantile regression technique to identify how differently real estate prices respond to a change in one unit of housing characteristic at different quantiles, without the need of accounting for spatial autocorrelation.

3. Model Specification

For the purpose of this study, the hedonic pricing model of residential real estate takes the following forms:

\[ P_i = f(H, N, \alpha, \beta). \]  

(2)
where \( P_i \) is the residential sales price of property \( i \); \( H_i \) is a vector of physical housing attributes associated with an apartment, \( N_i \) is a vector of neighborhood / locational variables, and \( \alpha \) and \( \beta \) are the estimated parameters associated with the exogenous variables.

A variety of econometric issues arises from estimating hedonic models, including the model specification, function form, the problems associated with heteroskedasticity, and spatial correlations. Ideally, model specification and function form should be determined by theoretical framework. Unfortunately, there is little theoretical guidance regarding model specification and restrictions imposed on function form, with an exception of the guidance in respect of the expected signs of certain coefficients associated with the variables. On one hand, model specification is largely determined by data availability and a \textit{priori} beliefs about the type of location and structural amenities that are relevant to each household. On the other hand, the choice of functional form is largely evaluated by empirical evidence. A typical approach is to compare the goodness of fit, Akaike information criterion (AIC) or Bayesian information criterion (BIC) from alternative functional forms, and then pick up the best fitting model.

In the ordinary least square (OLS) estimation, one of the classical assumptions is that the endogenous and residual variables are homoskedastic, which require the variances of error terms to be constant across observations. However, heteroskedasticity is often found to exist in cross-sectional or panel data due to the properties of the data. For example, larger or older dwelling units tend to have a larger error term than those of smaller or relatively new units. If this classical assumption is not held true, inaccurate standard errors and inefficient estimators are expected from the results. To test for the assumption of homoskedasticity, the White’s (1980) test
can be performed, which involves an auxiliary regression of the squared residuals on
the original regressors and their squares to test for the null hypothesis of no
heteroskedasticity against heteroskedasticity of some unknown general form. The test
statistic is computed by an auxiliary regression, where the squared residuals are
regressed on all possible (non-redundant) cross products of the regressors.

Following Koenecker and Hallock (2001) methodology, an alternative
methodology is the use of quantile regression which generalizes the concept of an
unconditional quantile to a quantile that is conditioned on one or more covariates. The
quantile can be defined through a simple alternative expedient as an optimization
problem. For example, the sample mean could be defined as the solution to the
problem of minimizing a sum of square residuals, and the median could be defined as
the solution to the problem of minimizing a sum of absolute residuals. The symmetry
of the piecewise linear absolute value function implies that the minimization of the
sum of absolute residuals must equate the number of positive and negative residuals.
Hence, it ensures that there are the same numbers of positive and negative
observations above and below the median. As the symmetry of the absolute value
yields the median, minimizing a sum of asymmetrically weighted absolute residuals
(i.e., simply giving differing weights to positive and negative residuals) would yield
the quantiles. Solving equation (3)

$$\min_{\xi \in \mathbb{R}} \sum \rho_{\tau}(y_i - \xi),$$

where the function $\rho_{\tau}(\cdot)$ is the tilted absolute value function that yields the $\tau$th sample
quantile as its solution. Least squares regression offers a model for how to define
conditional quantiles in an analogous fashion. If there is a random sample
\(\{y_1, y_2, \ldots, y_n\}\), we can solve it
\[
\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2 ,
\]

Then the sample mean, and an estimate of the unconditional population mean, \(EY\), can be obtained. If we replace the scalar \(\mu\) by a parametric function \(\mu(x, \beta)\) and solve
\[
\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mu(x_i, \beta))^2 ,
\]
we can then obtain an estimate of the conditional expectation function \(E(Y | x)\).

For quantile regression, we can simply go further to obtain an estimate of the conditional median function by replacing the scalar \(\xi\) in Equation (3) by the parametric function \(\xi(x_i, \beta)\) and setting \(\tau\) to 1/2. To obtain estimates of the other conditional quantile functions, we can replace the absolute values by \(\rho_\tau(.)\) and solve
\[
\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \xi(x_i, \beta)) .
\]

When \(\xi(x, \beta)\) is formulated as a linear function of parameters, the resulting minimization problem can then be solved very efficiently by linear programming methods.

The standard errors and confidence limits for the coefficient estimates can be obtained with asymptotic and bootstrapping methods. Both methods provide robust results (Koenkecker and Hallock 2001), with the bootstrap method considered more practical (Buchinsky, 1982; Efron, 1982; Hao and Naiman, 2007). Gould (1992; 1997) also suggest that the standard errors of coefficient estimates using the bootstrap method are significantly less sensitive to heteroskedasticity than the standard error estimates based on the method suggested by Rogers (1993).

---

2 See Hall (1992) for the detailed discussions on how to employ bootstrapping method to estimate the standard errors of the coefficient estimates.
4. Data Sources

To minimize the spatial effects upon residential property prices, a feasible approach is to select a sample of similar location-specific characteristics, relatively homogenous household tastes, and least variations in building design and quality such that the net effects of inherent attributes and location-specific factors tend to be similar. For the purpose of this study, we choose the City One Sha Tin for our case study because it comprises 10,642 small to-medium sized units in 52 residential blocks of different sizes and layouts, but with a relatively homogenous design. It is a standard mass housing estate located in the New Territories with a high trading volume at all times. Since the current study casts a focus on only one housing estate, the accessibility characteristics (such as accessibility to transport, amenities, and school, etc) and the external environment are more or less identical for all dwelling units of the estate.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Jarque-Bera</td>
</tr>
</tbody>
</table>

Data on housing prices, physical and location-specific characteristics are generated from the government official property transaction records compiled by a major real estate valuation firm. Observations with missing data for any of the variables described below are dropped from the analysis. This process yields a sample of 5,947 housing transactions. Real estate prices, $P$, represent the transaction price
(total consideration) of a residential property, which is recorded in HK dollars, inflation adjusted. \textit{GFA} represents the total gross floor area of a residential property, which is measured in square feet. \textit{AGE} represents the age of a residential property in years, which can be measured by the difference between the date of issue of the occupation permit and the date of transaction. \textit{FL} represents the floor level of a property in a residential building block. Apartment size, age and floor level are included as quadratic effects for the test of their nonlinear effect on prices. (See Table 1 for descriptive statistics.)

View is divided into three categories: building view, obstructive view and open view. It represents the type of view a property is facing respectively; for they equal 1 if a property is facing a particular view, 0 otherwise. The omitted category is open view so that coefficients may be interpreted relative to this category. The direction a property is facing is divided into 4 categories: \textit{NE}, \textit{SE}, \textit{SW} and \textit{NW}. It represents the direction a property is facing respectively; for they equal 1 if a property is facing a particular direction, 0 otherwise. The omitted category is \textit{NW} so that coefficients may be interpreted relative to this category. The car park, \textit{CP}, represents that a residential property transaction is associated with the sale of a car park. It is a dummy variable which equals 1 if the transaction is such a tie-in sale, 0 otherwise.

The financial variable is measured in real terms by using the “Monthly Price Indices for Selected Popular Private Domestic Developments” to deflate the series. This price deflator series is published by the Rating and Valuation Department, with the base year of 1999-2000=100. The indices are based on an analysis of price paid for apartments in selected housing developments, as recorded in their Sale and Purchase Agreement. Apart from the overall price indices for all residential properties within the selected housing estates, the indices are further broken down into price
indices for the small to-median sized properties and luxury properties, and can be further sub-divided by their price series in the urban areas and the New Territories respectively. Data are obtained from the Rating and Valuation Department.

Figure 1. City One Sha Tin. Source: www.maps.google.com

5. Empirical Results

Most analysis of hedonic pricing model has employed conventional least squares regression methods. However, it has been recognized that the resulting estimates of various effects on the conditional mean of real estate prices are not necessarily indicative of the size and nature of these effects on the lower tail of the price distribution. A more complete picture of covariate effects can be provided by estimating a family of conditional quantile functions. At any chosen quantile, one can ask how different are the corresponding real estate prices, given a specification of the
other conditioning variables. Table 2 presents a summary of the empirical results obtained by the traditional hedonic pricing model and the quantile regression. The estimated coefficient estimates for the linear regression and the 5th, 10th, 25th, 50th, 75th, 90th, 95th quantile regression coefficient estimates for property prices (along with their t-statistics), goodness of fit measures, and diagnostic statistics are shown. To correct for the observed heteroskedasticity and correlations among observations in cross-sectional data, this study employs HAC covariance to estimate the implicit prices of the housing attributes in the OLS specification. Most variables are statistically significant at conventional levels and have the expected signs.

The apartment size, age and floor level enter the model as quadratic effects. According to the linear regression model, while GFA tends to increase real estate prices up to the size of 1,257 square feet, it tends to decrease prices beyond 1,257 square feet. AGE tends to decrease prices up to 13.4 years, and increase prices beyond 13.4 years. FL tends to increase prices up to 24 floor level, and decrease prices after that threshold level. For quantile regression, the “optimal size” becomes bigger, with the exception of size at \( \tau = 0.05 \) and \( \tau = 0.1 \). At lower quantiles, such as at \( \tau = 0.1 \), it is about 1,212 square feet. At higher quantiles, it is about 1,348 square feet at \( \tau = 0.9 \), and 1,417 square feet at \( \tau = 0.95 \). For the “optimal age,” it is lower than the mean age at all ranges. At lower quantiles of \( \tau = 0.05 \) and \( \tau = 0.1 \), the “optimal floor level” is lower than the mean floor level. At higher quantiles, the optimal floor levels are 26, 27 and 33 at \( \tau = 0.75 \), 0.9 and 0.95 respectively, all of which are greater than the mean floor level.

Homebuyers generally do not favor properties that have a building or obstructive view, for most of them prefers properties with an open view, green view, or sea view. Empirical results demonstrate that homebuyers of higher-priced properties are more
concerned about the type of view their properties have, and they are not willing to opt for properties with a building or obstructive view unless a bigger discount is offered to them than to the homebuyers of lower-priced properties. This phenomenon is represented by bigger and negative estimated coefficients of these two variables at higher quantiles than those of their mean values and the lower quantiles.

An apartment with a car park obviously commands a greater price premium than an apartment without one, about HK$167,000, according to the ordinary least squares estimates of the mean effect, but is clear from the quantile regression results that the disparity is much larger in the lower quantiles of the distribution and considerably smaller in the higher tail of the distribution. For example, a car park commands HK$314,000 at 0.05 quantile, but only costs about HK$52,000 at 0.95 quantile. The least squares estimate of the mean car park effect is thus mainly captured by the lower tail of conditional distribution. The conventional least squares confidence interval does a poor job of representing this range of disparities.
Table 2. Quantile Regression Coefficient Estimates

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>0.05</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-819906.8*</td>
<td>-908418.7*</td>
<td>-834659.3*</td>
<td>-733111.0*</td>
<td>-727671.7*</td>
<td>-864617.1*</td>
<td>-966063.9*</td>
<td>-898253.3*</td>
</tr>
<tr>
<td>GFA</td>
<td>6980.7*</td>
<td>6341.4*</td>
<td>6344.7*</td>
<td>6276.1*</td>
<td>6533.5*</td>
<td>7121.0*</td>
<td>7253.0*</td>
<td>7105.9*</td>
</tr>
<tr>
<td>GFA^2</td>
<td>-2.8*</td>
<td>-2.9*</td>
<td>-2.6*</td>
<td>-2.3*</td>
<td>-2.3*</td>
<td>-2.3*</td>
<td>-2.7*</td>
<td>-2.5*</td>
</tr>
<tr>
<td>AGE</td>
<td>-62374.2*</td>
<td>-20258.6**</td>
<td>-83106.3*</td>
<td>-50450.0*</td>
<td>-58956.6*</td>
<td>-58813.0*</td>
<td>-45180.4*</td>
<td>-43096.0*</td>
</tr>
<tr>
<td>AGE^2</td>
<td>2328.3*</td>
<td>906.3*</td>
<td>1518.4*</td>
<td>1946.8*</td>
<td>2295.9*</td>
<td>2353.6*</td>
<td>1990.0*</td>
<td>1945.8*</td>
</tr>
<tr>
<td>FL</td>
<td>16422.8*</td>
<td>13850.3*</td>
<td>15519.0*</td>
<td>13997.9*</td>
<td>14189.8*</td>
<td>13161.8*</td>
<td>12579.1*</td>
<td>11144.5*</td>
</tr>
<tr>
<td>FL^2</td>
<td>-336.8*</td>
<td>-314.4*</td>
<td>-341.0*</td>
<td>-291.1*</td>
<td>-291.9*</td>
<td>-248.5*</td>
<td>-227.0*</td>
<td>-166.1*</td>
</tr>
<tr>
<td>BUILDING</td>
<td>-21157.5**</td>
<td>-23282.9</td>
<td>-21527.4**</td>
<td>-16268.3**</td>
<td>-27095.4*</td>
<td>-33604.4*</td>
<td>-56547.2*</td>
<td>-62955.8*</td>
</tr>
<tr>
<td>OBSTRUCTIVE</td>
<td>-16767.3***</td>
<td>-8244.9</td>
<td>-3968.2</td>
<td>-8044.6</td>
<td>-18585.7**</td>
<td>-28704.9*</td>
<td>-57754.1*</td>
<td>-83214.3*</td>
</tr>
<tr>
<td>NE</td>
<td>73641.3*</td>
<td>58639.2*</td>
<td>65986.5*</td>
<td>66531.3*</td>
<td>65990.8*</td>
<td>74192.7*</td>
<td>77399.8*</td>
<td>60520.2*</td>
</tr>
<tr>
<td>SE</td>
<td>7654.7</td>
<td>2617.3</td>
<td>5470.1</td>
<td>5632.0</td>
<td>5705.4</td>
<td>9030.7</td>
<td>19649.5</td>
<td>3858.6</td>
</tr>
<tr>
<td>SW</td>
<td>82157.7*</td>
<td>54643.3*</td>
<td>64184.5*</td>
<td>66598.0*</td>
<td>70590.7*</td>
<td>81316.6*</td>
<td>92479.1*</td>
<td>84518.4*</td>
</tr>
<tr>
<td>CP</td>
<td>166604.6*</td>
<td>313664.0*</td>
<td>281882.5*</td>
<td>238394.7*</td>
<td>195772.4*</td>
<td>141278.8*</td>
<td>95458.6*</td>
<td>51636.1*</td>
</tr>
<tr>
<td>R^2</td>
<td>0.9248</td>
<td>0.3955</td>
<td>0.5017</td>
<td>0.6283</td>
<td>0.7343</td>
<td>0.7994</td>
<td>0.8149</td>
<td>0.8332</td>
</tr>
</tbody>
</table>

Notes: * indicates statistically significant at 1 percent confidence level; ** indicates statistically significant at 5 percent confidence level. *** indicates statistically significant at 10 percent confidence level. The R^2 for quantile regression is the pseudo R^2.

Table 3. Optimal Level of Housing Characteristics

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>0.05</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFA</td>
<td>1257.3</td>
<td>1096.2</td>
<td>1212.2</td>
<td>1375.4</td>
<td>1390.8</td>
<td>1318.1</td>
<td>1347.6</td>
<td>1416.9</td>
</tr>
<tr>
<td>AGE</td>
<td>13.4</td>
<td>11.2</td>
<td>12.5</td>
<td>13.0</td>
<td>12.8</td>
<td>12.5</td>
<td>11.4</td>
<td>11.1</td>
</tr>
<tr>
<td>FL</td>
<td>24.4</td>
<td>22.0</td>
<td>22.8</td>
<td>24.0</td>
<td>24.3</td>
<td>26.5</td>
<td>27.7</td>
<td>33.6</td>
</tr>
</tbody>
</table>
6. Concluding Remarks

The objective of this paper is to investigate how differently homebuyers value specific housing characteristics across different quantiles of conditional distribution. Although linear regression estimates the mean value of the response variable for given levels of the predictor variables, the results are quite different from the specific data points within the sample, depending on which side of the distribution those particular points of interest lie. Particularly, the quantile regression parameter estimates the change in a specified quantile of the response variable produced by a one-unit change in the predictor variable. This allows for a comparison of how specific percentiles of real estate prices may be more affected by certain properties’ characteristics than other percentiles. This is reflected in the change in the size of the regression coefficient.

The distinction between linear and quantile regression is best explained by Mosteller and Tukey (1978) that “[w]hat the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of x’s. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a corresponding incomplete picture for a set of distributions.”

Empirical results suggest that homebuyers’ tastes and preferences for specific housing attributes vary greatly across different quantiles of conditional distribution. This is simply due to the fact that individual’s tastes and preferences are unique such that some homebuyers place a higher valuation on certain housing characteristics than others. The quantile regression approach complements the least squares by identifying how differently real estate prices response to a change in a one-unit of housing
characteristic at different quantiles, rather than estimating the constant regression coefficient representing the change in the response variable produced by a one-unit change in the predictor variable associated with that coefficient. It allows buyers of higher-priced properties to behave differently from buyers of lower-priced properties even if they are within one single housing estate, which better explains the real world phenomenon.

References


Buchinsky M, 1982, “Methodological issues in quantile regression” mimeo


Hall P, 1992, *The bootstrap and Edgeworth expansion* (Springer-Verlag, New York; Berlin)


Rating and Valuation Department, 1981-2005, *Hong Kong Property Review* (Government of the Hong Kong Special Administrative Region, People’s Republic of China)


