THREE ESSAYS ON THE APPLICATION OF QUANTILE REGRESSION IN REAL ESTATE ECONOMICS

By

Yixiu Zhou

Department of Economics and Finance
Middle Tennessee State University
Murfreesboro, TN 37129
yz2e@mtsu.edu
zhou.yixiu@gmail.com

December 2008

JEL Codes: R31, C21, C29
Keywords: Quantile regression, housing price, flooring, time-on-market

APPROVAL PAGE

THREE ESSAYS ON THE APPLICATION OF QUANTILE REGRESSION IN REAL ESTATE ECONOMICS

BY

YIXIU ZHOU

A DOCTORAL DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY / ECONOMICS

MIDDLE TENNESEE STATE UNIVERSITY

DECEMBER, 2008

APPROVED BY: Joachim Zietz

Dr. Joachim Zietz, Committee Chair

David A. Penn

Dr. David Penn, Committee Member

E. Anthon Eff

Dr. E. Anthon Eff, Committee Member

Gregory E. Givens

Dr. Gregory E. Givens, Director, Economics Graduate Program

Charles L. Baum

Dr. Charles L. Baum, Department Chair, Economics & Finance

Michael D. Allen

Dr. Michael D. Allen, Dean of College of Graduate Studies

ABSTRACT

This dissertation is composed of three individual essays on the use of conditional quantile regression in real estate economics. The first essay discusses some limitations of the traditional conditional quantile regression methodology. A modification is proposed to improve the interpretability of conditional quantile regression estimates for applications to hedonic price functions in real estate valuation. The second essay provides empirical applications of the methodology suggested in the first essay to analyze the implicit prices of different types of flooring in single-family homes. The third essay suggests that conditional quantile regression can be a viable alternative to duration models to analyze the determinants of the length of time between the market listing of a home and its sale. The essays employ variants of a recently suggested spatial-temporal technique to identify neighborhood effects to avoid spatial autocorrelation and endogeneity problems. The essays rely on data from the Multiple Listing Service (MLS) for single-family homes in Rutherford County, TN, and the county planning commission. The data cover the years 2003 to 2007.

The first essay illustrates that the traditional quantile regression estimates are likely to overestimate the coefficient dispersion across quantiles. As a direct consequence, hedonic price functions in real estate applications may underestimate the prices of homes at the lower-end of the distribution and overestimate prices at the upper end. Unconditional quantile regression is shown to suffer from the same problem, except to a larger degree. An adjustment factor is proposed for the traditional conditional quantile regression estimates to minimize the prediction error.

The second essay applies the methodology proposed in the first one to examine the role of different types of flooring in determining house prices. The results suggest that there are large differences in the implicit values attached to different types of flooring. Almost uniformly across the sales price distribution, finished wood is the most valued flooring type, closely followed by marble and tile. Carpet is the standard flooring type used by almost 96 percent of all homes and, therefore, does not add extra value. The use of vinyl flooring tends to lower the value of a house across all price ranges. The essay identifies large differences in implicit prices across quantiles for different combinations of flooring types. For example, a combination of finished wood, tile, and parquet adds the most value to lower and medium priced homes, while a combination of carpet, wood, and marble has a particularly high implicit value at the upper end of the price distribution.

The third essay shows that conditional quantile regression is a viable technique to study the determinants of the time it takes from the listing of a home to its sale (TOM). Quantile regression has a number of advantages over the traditional methods of duration analysis, such as the Cox proportional hazards model. It allows the determinants of TOM to affect the time on the market differently across the distribution of TOM and the results are much easier to interpret, which is an important consideration for practical applications. For the data set at hand, significant differences are found across the distribution of TOM for a number of variables. For example, a high list price relative to sales price of neighboring properties prolongs TOM perceptively for houses that sell quickly. The same applies to houses which are located in a less desirable school zone.

Table of Contents

List of Tables

Essay 1. A Modified Conditional Quantile Estimator: An Application to Hedonic Price Functions in Real Estate Economics

Essay 2. Types of Flooring As a Determinant of The Sale Price of Single Family Homes: A Quantile Regression Approach

Essay 3. An Empirical Study of Time-On-Market Using Quantile Regression

List of Figures

Essay 1. A Modified Conditional Quantile Estimator: An Application to Hedonic Price Functions in Real Estate Economics

Essay 2. Types of Flooring As a Determinant of The Sales Price of Single Family Homes: A Quantile Regression Approach

Essay 3. An Empirical Study of Time-On-Market Using Quantile Regression

Overview

Quantile regression is a robust, flexible, and often more informative estimation technique than classical least squares regression. By minimizing the sum of asymmetrically weighted absolute deviations it is robust to outliers, which can perceptively affect least squares estimates. Such robustness is important in real estate applications, where a researcher often deals with large data sets that tend to contain outliers even after attempts to remove apparently unreasonable observations. Quantile regression is also not subject to some of the estimation and interpretation problems that arise when such crucial assumptions for least squares as homoskedasticity are not met. This is because quantile regression estimates are not subject to the tight distributional assumptions of least squares.

Quantile regression is more flexible and more informative than least squares because it allows for coefficients to vary across the quantiles of sales price or some other dimension of interest, such as time on the market. Varying coefficients allow the data as opposed to the researcher to decide whether an explanatory variable has the same impact on the dependent variable across all points of the distribution of the dependent variable. This flexibility is important for applications in real estate valuation because houses are heterogeneous in many dimensions but most importantly in terms of price. The price of a house can in effect be thought of as a natural segmentation variable because expensive homes are not affordable to poor households and cheap homes are not desirable for affluent households. Assuming that the demand for and supply of housing attributes are different across the price spectrum, one can make a strong case that it is sensible to allow for different implicit prices across the price distribution. Quantile regression is ideally suited to implement this idea because it can generate a different set of coefficients for each quantile without truncating the sample by price and, therefore, causing a severe sample selection bias. All observations are incorporated for each quantile regression; the observations are merely weighted differently from quantile to quantile.

Quantile regression gives the researcher the flexibility to select a conditional quantile of the response variable as the dependent variable. This is of significant practical importance because real estate professionals often have a particular interest in only one or two segments of the distribution. For instance, realtors only selling expensive homes require a different information set than those typically selling inexpensive homes.

Quantile regression has been introduced into real estate economics only recently. Therefore, a number of practical questions still await an answer. One such question is how to deal with the problem that the conditional quantile estimator does not predict the unconditional quantile points very accurately. The first essay demonstrates this point and examines whether a possible solution lies in the use of a technique that has recently been suggested for applications of quantile regression in other areas of economics. It is found that the so-called unconditional quantile estimator does not solve the problem faced in the context of real estate valuation. In fact, it exacerbates it. The first essay introduces as an alternative a simple adjustment method and demonstrates its usefulness. The second essay provides a fully fledged application of the adjustment methodology for a practical issue in real estate valuation: identifying the implicit prices of different types of flooring. The third essay introduces quantile regression as an alternative to duration analysis in real estate economics. So far, the real estate literature on the time it takes to affect a sale or similar applications has been focused on duration models, such as the Cox proportional hazards model. Like least squares, the proportional hazards model assumes fixed parameters for all observations. In addition, the method is difficult to use and interpret, something which makes it less attractive for practical applications. The third essay demonstrates that by allowing parameters to differ along the distribution of the time it takes to sell a house quantile regression can uncover valuable information of relevance to speed up the sale of a house. The methodology as well as the results should be of significant interest to real estate practitioners.

A Modified Conditional Quantile Estimator: An Application to Hedonic Price Functions in Real Estate Economics

Abstract

The paper suggests a modification of the conditional quantile estimator to effectively remove the key problem of using this estimator to derive the implicit valuations of housing characteristics via hedonic price functions: its inability to predict the unconditional quantile values of the dependent variable. The suggested modification of the conditional quantile estimator involves a simple adjustment factor between zero and one, which is particular to a given data set and can be derived with little effort. How the suggested methodology can be applied in practice is illustrated on a data set consisting of transactions data on nearly 5,000 home sales for the years 2006 and 2007.

1 Introduction

The quantile regression methodology, as first introduced by Koenker and Basset (1978), has been the subject of much recent methodological research and is now used in many fields inside and outside of economics and finance. [1] Examples of recent methodological innovations are the contributions by Peng and Huang (2008) on survival analysis and by Chernozhukov and Hansen (2008) and Horowitz and Lee (2007) on instrumental variable estimation.

Quantile regression has been applied extensively in economics, for example, in labor economics to analyze wage inequality and wage structure (e.g., Martins and Pereira 2004), in population economics to study fertility (e.g., Miranda 2008), in health economics to consider health care reform (Winkelmann 2006), or in education economics to derive the impact of education on the labor market (e.g., McGuinness and Bennett 2007) or the consequences of school composition on educational outcomes (e.g., Rangvid 2007). In international economics, quantile regression has been employed to analyze the behavior of exchange rates (Nikolaou 2008) or international linkages at the plant level (Yasar and Morrison 2007), and in development economics it has been used to analyze urban-rural inequality (Nguyen et al. 2007).

In financial research, quantile regression is now a common method to value the market risk (Taylor 2008 and Bassett and Chen 2001) and to analyze the determinants and consequences of capital structure (Fattouh et al. 2008, Margaritis and Psillaki 2007)

[1] An overview of early applications of quantile regression is provided in Koenker and Hallock (2001). In the last few years, the number of applications of quantile regression has increased sharply. The electronic version of the *Journal of Economic Literature* (EconLit) now lists hundreds of studies using or modifying quantile regression.

or of stock repurchases (Billett and Xue 2007). Zietz et al. (2008) employ quantile regression to estimate hedonic price functions for use in real estate valuation.

However, Firpo et al. (2007) have cast some doubt on the practical usefulness of the traditional quantile methodology. The authors argue that the results of the traditional conditional quantile methodology cannot generally be interpreted in analogy to those of OLS regressions. This applies, in particular, to the interpretation of the estimated coefficients as the marginal effects of changes in the regressor variables on the unconditional quantiles of the dependent variable. Given the insights provided by Firpo et al. (2007), the purpose of this paper is (a) to lay out, for a concrete application of conditional quantile regression to the valuation of residential housing, the interpretation problem that arises because the conditional quantile predictions do not equal the unconditional quantile points of the dependent variable, (b) to suggest a simple adjustment to the conditional quantile estimator to circumvent this interpretation problem, and (c) to provide some concrete guidance on applying the suggested methodology for estimating hedonic price functions in real estate valuation.

The study makes use of a data set consisting of approximately 5,000 transactions of residential homes for the years 2006 and 2007. Sales prices and home characteristics are taken from a multiple listing service. They are supplemented by various location variables and other local government data.

The remainder of this paper is organized as follows. Section 2 presents the traditional quantile methodology, the interpretation problems that arise, and possible solutions to these problems. Section 3 contains a brief description of the data. Section 4 discusses the empirical estimation results. Section 5 concludes.

2 The Quantile Methodology

2.1 An Interpretation Problem of Conditional Quantile Regression

The unconditional mean of variable y is given by the value of $\overline{y}$ in the minimization problem

$$\min_{\overline{y}} \sum_i \left(y_i - \overline{y}\right)^2 . \tag{1}$$

By analogy, the unconditional quantile q ($0 < q < 1$) of variable y can be found as the value of y_q in the minimization problem

$$\min_{y_q} \sum_i \left|y_i - y_q\right| h_i , \tag{2}$$

where

$$h_i = \begin{cases} 2q \\ 2-2q \end{cases} \text{if} \quad \begin{cases} y_i - y_q > 0 \\ y_i - y_q \leq 0 \end{cases} \tag{3}$$

For a value of $q = 0.5$, the value of y_q will equal the median of series y.

Ordinary least squares (OLS) and conditional quantile regression (CQR) can be thought of as natural extensions of equations (1) and (2) in that the scalars $\overline{y}$ and y_q are replaced by the predictor variable m_i that depends on $k + 1$ regressor variables (x_j) in the form

$$m_i = \sum_{j=0}^{k} b_j x_{j,i} , \tag{4}$$

where the b_j identify the coefficients to be estimated.[2] Rather than minimizing with respect to the scalars $\overline{y}$ or y_q, OLS and CQR minimize with respect to the set of coefficients $\{b_j\}_{j=0}^{k}$. In this way, both OLS and CQR provide predictions that are conditional on the set of regressors $\{x_j\}_{j=0}^{k}$.

The unique characteristic of OLS is that its prediction conditional on the estimated coefficients and the means of the regressors is equal to the observed or unconditional mean of the dependent variable. This very convenient characteristic of OLS can be written in equation form as

$$\mathrm{E}\mathbf{y} = \mathrm{E}\left[\mathrm{E}\left(\mathbf{y}|\mathbf{x}\right)\right] = \mathrm{E}[\mathbf{m}] = \mathrm{E}[\mathbf{xb}] = \mathbf{b}'\mathrm{E}[\mathbf{x}] \,, \tag{5}$$

where **y**, **x**, **m**, and **b** are the appropriately dimensioned vector/matrix equivalents of y_i, x_j, m_i, and b_j. This property allows interpreting the estimated coefficient of the j-th regressor $\left(\hat{b}_j\right)$ as the response of the average value of the dependent variable $\left(\overline{y}\right)$ to a unit change in the average value of regressor x_j, $\left(\overline{x}_j\right)$,

$$\hat{b}_j = \frac{\partial \overline{y}}{\partial \overline{x}_j}. \tag{6}$$

This key characteristic does not transfer to the CQR of the q-th quantile,

$$\hat{b}_{j,q} \neq \frac{\partial y_q}{\partial \overline{x}_j}, \tag{7}$$

[2] Equation (4) allows for nonlinear functional forms in that the j regressor variables can contain higher orders or logarithms of continuous variables.

where $\hat{b}_{j,q}$ is the estimated coefficient of the j-th regressor for the q-th quantile of the dependent variable. For CQR, the unconditional q-th quantile value of y (y_q in equation 2) is not in general equal to the conditional q-th quantile of y evaluated at the mean values of the regressors,

$$y_q \neq \sum_j \hat{b}_{j,q} \bar{x}_j. \tag{8}$$

Figure 1 illustrates the above points. It is based on the data set used for this study, which is discussed later in some detail. The figure presents, for each quantile from 0.1 to 0.9, (a) the mean value of the dependent variable $(\bar{y})$, (b) the unconditional quantile points (y_q), and (c) the CQR predictions evaluated at the mean values of the regressor variables $\left(\hat{y}_{q,cqr} = \sum_{j=0}^{k} \hat{b}_{j,q} \bar{x}_{j,q}\right)$.[3] The quantile values predicted by the CQRs lie somewhere between the mean of the dependent variable $(\bar{y})$ and the observed unconditional quantile values (y_q). The conditional quantile predictions take into account the movement of the observed quantile values away from the median, but their predictive power decreases with increasing distance from the median. The extent of the observed price dispersion is not captured by the predictions of the conditional quantile regression when the estimated coefficients are evaluated at their sample means. The conditional quantile estimate $\hat{b}_{j,q}$ can be interpreted as the marginal effect of a change in the average value of x_j $(\bar{x}_j)$ on the conditional quantile of y but not, in general, on its

[3] The quantile points are connected by straight line segments.

unconditional quantile (y_q in equation 2). Therefore, conditional quantile regression is not appropriate for research questions that examine changes in the unconditional distribution of the dependent variable to changes in the mean value of one or more of the regressors. An example from real estate valuation would be the question whether a change in average industrial pollution affects the prices of homes at the unconditional 90th quantile more than at the unconditional 10th quantile and, therefore, changes the price dispersion between low- and high-priced properties.

2.2 Alternative Strategies to Address the Interpretation Problem

There is no straightforward solution to dealing with the issue that derives from the fact that the predicted conditional quantile values, when evaluated at the mean values of the regressors, are not equal to the corresponding unconditional quantile values. One strategy to deal with the apparent interpretation issue is to discard conditional quantile regression altogether and move from conditional quantile regression, as introduced by Koenker and Bassett (1978), to unconditional quantile regression as recently suggested by Firpo et al. (2007). An alternative strategy would be to check to what extent the conditional quantile estimator can be adapted or reinterpreted to provide meaningful predictions of the unconditional quantiles. Both strategies will be considered in turn.

2.2.1 Unconditional Quantile Regression

In its simplest form,[4] an unconditional quantile regression (UQR) involves solving equation 9 for vector $\boldsymbol{\beta}$,

$$\min_{\boldsymbol{\beta}} \sum_{i} \left[r_i - m_i\left(\mathbf{x}_i, \boldsymbol{\beta}\right) \right]^2, \tag{9}$$

where both $\boldsymbol{\beta}$ and r_i are specific to a given quantile. It is apparent that equation 9 represents an OLS estimator, but with the twist that the dependent variable is r_i rather than y_i. The variable r_i is a transformation of a 0/1 indicator variable that is particular to the chosen quantile. In brief, r_i is given as

$$r_i = c_1 tq_i + c_2 \tag{10}$$

where tq is a 0/1 indicator variable that is unity for all values of the dependent variable y_i that are larger than the value of the dependent variable at the chosen quantile q $\left(y_q\right)$. Parameter c_1 is derived as

$$c_1 = \frac{1}{f_{y_q}}, \tag{11}$$

where f_{y_q} is the value of the Gaussian kernel density function with optimal bandwidth that corresponds to y_q. Parameter c_2 is calculated as

$$c_2 = y_q - c_1(1-q). \tag{12}$$

The unconditional quantile estimator of equation 9 has the property that the predicted value of m_i for quantile q evaluated at the mean of the regressor variables x_j

---

[4] As an alternative, a logit model can be estimated, with the 0/1 variable tq_i on the left. The marginal effects of the unconditional quantile estimator would then be calculated as the product of the parameter c_1 and the marginal effects of the logit regression with respect to the independent variables.

$\left(\bar{x}_j\right)$ is equal to the unconditional quantile value of y, which is identified as y_q in equation 2 and in Figure 1,

$$y_q = \sum_{j=0}^{k} \hat{\beta}_{j,q} \bar{x}_j. \tag{13}$$

Hence, parameter vector $\boldsymbol{\beta}$, as estimated from equation 9 for a particular quantile q, maps changes in the average values of the elements of $\mathbf{x}$ into changes of the unconditional quantile q of dependent variable y. This is analogous to OLS predicting changes in the mean of the dependent variable from changes in the mean values of the right-hand side variables.

UQR is helpful as a technique in situations where it is of interest to trace changes in the unconditional quantile points to changes in the mean values of one or more of the independent variables. That, for example, may be the case if an answer is sought to the question whether a decrease in the average crime rate or in the average concentration of pollutants raises the prices of expensive homes more than those of inexpensive ones. However, by force of its focus on changes in the mean value of the regressors, UQR is not helpful for predicting changes in the unconditional quantile points if there is little or no change in the sample means of the independent variables of interest, yet sizable but largely offsetting changes in the quantiles.

UQR is also not helpful for practical valuation applications involving hedonic price functions because the perspective of most valuation questions is different from that of UQR. In real estate valuation, the primary focus is on mapping *individual* characteristics of real estate objects into market prices not on predicting the consequences for price of changes in the mean values of the regressors. At issue is, for example, how to value a particular property given its characteristics or how to assess the implicit market value of a particular property feature, such as a pool. Using UQR-estimated parameters to map the features of a particular property into a market price would be an inappropriate use of UQR. The market price would be considerably overestimated by UQR if the values of the characteristics happen to be above their sample means, and vice versa. UQR correctly maps only sample means and their changes into market prices, not values of characteristics possibly far away from their sample means. Figure 2 illustrates this point. It provides, at intervals of 0.1, (a) the observed unconditional quantile values of the dependent variable $\left(y_q\right)$ (b) predictions of the unconditional quantile values based on the parameter estimates of the corresponding UQR and smoothed regressor values [5] $\left(\hat{y}_{q,uqr} = \sum_{j=0}^{k} \hat{\beta}_{j,q}\hat{x}_{j,q}\right)$, and (c) the corresponding predictions based on the CQR estimates $\left(\hat{y}_{q,cqr} = \sum_{j=0}^{k} \hat{b}_{j,q}\hat{x}_{j,q}\right)$.

2.2.2 A Modified Conditional Quantile Estimator

Figure 2 not only reveals that UQR is ill adapted at predicting the dependent variable based on values of the regressor variables other than their means. Figure 2 also suggests that CQR does not perform well either at mapping observed individual characteristics into observed market prices. The CQR predictions are much better than those based on the UQR estimates, but the prediction bias is in the same direction as the

[5] How the smoothed regressor values $\left(\hat{x}_{j,q}\right)$ are derived is discussed in detail in the data section. In brief, they are those regressor values that are typical or predictable for the given quantile points of the dependent variable.

one for UQR-based predictions: the price dispersion between lower- and upper-priced segments is overstated.

Figure 2 has important implications for the practical use of CQR. It appears that CQR has little to recommend itself for pricing properties to market because the unconditional quantiles are not predicted well. In addition, if the predictions are consistently biased, little confidence can be attached to interpreting the estimated parameter values as implicit prices of a property's characteristics.

To make CQR useful for valuation purposes, the key task is to improve CQR's ability to predict the unconditional quantile points of the dependent variable. Figures 1 and 2 provide some helpful hints at how this task can be accomplished. In particular, Figure 1 shows that the CQR coefficient estimates evaluated at the mean values of the regressors significantly underestimate the actual price dispersion of the unconditional quantile points across quantiles. Figure 2, by contrast, suggests that the price dispersion is systematically overstated for the regressor values that are typically associated with the particular quantiles. It is apparent that the overestimate of the price dispersion in Figure 2 is significantly less than the underestimate of the price dispersion in Figure 1. The conclusion is that the CQR coefficients need to be evaluated at a weighted average of (a) the sample means of the regressors and (b) those values which the regressors typically take on at the different quantiles. In equation format, this conclusion can be written as

$$y_q \cong \sum_{j=0}^{k} \hat{b}_{j,q} \left[w\overline{x}_j + (1-w)\hat{x}_{j,q} \right], \tag{14}$$

where w $(0 \le w \le 1)$ is the weight attached to the sample means of the regressor variables and where $\hat{x}_{j,q}$ is the value of regressor variable j that is typically associated with the q-th unconditional quantile of y.

Figure 3 depicts the CQR-based predictions of equation 14 for various values of the weight w. As in Figure 2, the objective is to predict the unconditional quantile values (y_q). For $w=1$, equation 14 replicates the line identified as $\hat{y}_{q,cqr}$ in Figure 1. For $w=0$, equation 14 replicates $\hat{y}_{q,cqr}$ in Figure 2. If $w=0.2$, equation 14 approximates the unconditional quantile values (y_q). This approximation of the unconditional quantiles is not as perfect as the one that is achieved when the UQR estimates are evaluated at the mean values of the regressors. For example, Figure 3 reveals some overestimate of the unconditional quantile values at quantile points 0.4 and 0.9. However, the predictions with weight $w=0.2$ are significantly better than those that can be achieved by standard CQR and weight $w=0$.

The modification of the CQR estimator suggested by equation 14 is simple to implement but rather effective at allowing CQR-based estimates to predict the unconditional quantiles of the independent variable. Unlike the UQR estimator, equation 14 is responsive to changes not only in the mean values of the regressors but also in the regressor values that are typically associated with each quantile. In other words, equation 14 will predict a change in a particular unconditional quantile point even if there is only a change in the regressor values at that quantile point but no change in the regressor means, due to an offsetting change in a different quantile point.

2.3 Applying the Modified Conditional Quantile Estimator

Assume that a sample of property prices and associated property characteristics are available to estimate traditional conditional quantile regressions at intervals of 0.1 or at a finer detail. Also assume that an unsold property i needs to be priced to market and the implicit values of its characteristics determined. The steps that are needed to accomplish this task provide a simple demonstration of how to use the modified conditional quantile estimation given by equation 14 in practice.

Since there is no value for the dependent variable of property i, it is not clear which CQR to apply in this case. However, the estimates of either OLS or CQR for quantile 0.5 can be used to price the property to market because the unconditional quantile points of the predicted values of both OLS and CQR for quantile 0.5 closely match the unconditional quantile points of the dependent variable. No further analysis is needed if a market price is all that is needed for property i. In particular, there is no need to use quantile regression. This conclusion changes if there is a need to identify implicit prices for the characteristics of property i and the predicted market price is not identical to the mean or median of the sample distribution of prices.

The first step in identifying these implicit prices is to determine the unconditional quantile point qi that matches property i's predicted market price. The second step consists of estimating a standard CQR for quantile point qi. Assume the coefficients of this standard CQR are given as $\hat{b}_{j,qi}, j = 0,..,k$. In step three, the implicit prices are derived from these CQR estimates according to equation 14. Because by equation 14 we have

$$y_{qi} \cong \sum_{j=0}^{k} \hat{b}_{j,qi}\left[w\overline{x}_j + (1-w)x_{j,i}\right], \tag{15}$$

where the weight w is assumed to be known for the given data set and where $x_{j,i}$ is the value of characteristic j of property i that needs to be priced, the implicit price of characteristic j is given as the partial derivative,

$$\left.\frac{\partial y_{qi}}{\partial x_{j,i}}\right|_{d\overline{x}_j=0} \cong \hat{b}_{j,qi}(1-w). \tag{16}$$

According to equation 16, the standard CQR estimate $\left(\hat{b}_{j,qi}\right)$ overestimates the true quantile coefficient by the factor $(1-w)$. This is the factor by which standard conditional quantile estimates need to be adjusted to ensure that the predictions of CQRs approximate the associated unconditional quantile values.

Equation 16 assumes that the mean value of variable x_j is not affected by marginal changes in the value of $x_{j,i}$. This cannot reasonably be assumed for all variables. It is certainly not the case for the regression constant. However, it also does not apply to binary variables that are by definition equal to zero or unity for the complete sample if they are assumed to be zero or unity for a particular quantile. This is relevant, for example, for 0/1 indicator variables that capture the rise in general inflation from one year to the next or for seasonal dummy variables. In these cases, the adjustment factor in equation 16 needs to be set at zero because the variables $\overline{x}_j$ and $x_{j,i}$ in equation 15 would move by the same amount.

3 Data

The data set consists of 4,990 one-family house transactions in Rutherford County, Tennessee, recorded in 2006 and 2007. In line with the literature, the sales price is assumed to be affected by a number of factors, including the structural attributes of the house and its location. The values attached to these various house characteristics are captured by a hedonic equation

$$\ln P = f(S, L),$$

where P is the selling price, S a vector of structural attributes, and L location variables. Table 1 provides an overview of the variables and their definitions. Table 2 summarizes the basic statistics of the variables.

The issue of spatial autocorrelation receives particular attention in this study. It is captured in the spatial lag variable *splag*. There is a large literature on spatial effects and how they can be considered in deriving parameter estimates. Key references are Anselin (1988, 2001) and Pace et al. (2000). There is a dearth of literature on incorporating spatial effects into a quantile regression framework. Recent work by Zietz et al. (2008) involves an instrumental variables estimator based on a spatial lag model that adapts the two-stage quantile method developed by Kim and Muller (2004). This estimator, however, is time consuming to implement as it involves bootstrapping in the context of a two-stage estimation process. The complexity of the estimation process results from the fact that a standard spatial lag approach is used, where the spatial lag variable is constructed as the product of a $n \times n$ weight matrix W and the $n \times 1$ dependent variable vector y. By construction, this spatial lag variable is endogenous. The estimation process could be simplified greatly if the spatial lag variable were not endogenous. The method suggested by Pace et al. (2000) manages to accomplish just this objective. The key point of this method is to enter only those properties into the construction of the spatial lag variable that sold prior to the date that the property at observation i sold. This avoids the endogeneity problem and two-stage estimation methods.

To make the spatial lag variable more meaningful as a neighborhood index, the present paper first identifies the ten nearest neighbors that sold prior to property i.[6] Second, only those properties with a distance of less than or equal to 0.3 miles are retained for the construction of the spatial lag variable. In practice, this last condition reduces the number of neighbors below ten for almost every observation.

Smoothing has the objective to find representative values of the independent variables that are matched to the given conditional quantiles of the dependent variable. These are identified as $\hat{x}_{j,q}$ in the previous section. Smoothed regressor variables are obtained by adapting the inverse regression idea. More specifically, for each regressor variable x_j, an auxiliary regression is run on the complete sample with the regressor variable on the left and the regressand (y) on the right,

$$x_j = g(y). \tag{17}$$

The value predicted by this auxiliary regression for y_q is taken to be the smoothed regressor value that corresponds to the unconditional quantile point q of the dependent variable (y_q),

$$\hat{x}_{j,q} = g(y_q). \tag{18}$$

[6] The data cover the years 2006 and 2007. The data for the year 2005 are used to construct the spatial lag variable for those properties that sold in 2006. The data for 2005 and 2006 are utilized to construct the spatial lag for those properties that sold in 2007.

Depending on the type of regressor, different auxiliary regressions are used. For a continuous regressor variable, a *lowess* regression with standard bandwidth (0.8) is employed. A *lowess* regression has the advantage over OLS that it is able to capture without manual intervention a nonlinear relationship that may exist between regressor and independent variable. The smoothing of dichotomous regressor variables uses the predictions of a logit regression. Predictions from poisson regressions are employed for those variables that represent count data, such as the number of bedrooms.

4 Estimation Results

All reported regression estimates use the log of the sales price as the dependent variable. The right-hand side variables enter in linear form. Consequently, all estimated coefficients can be interpreted as the approximate percentage change in price that results from a small change in the right-hand side variable.[7]

The estimates of the CQR and UQR coefficients are presented in Tables 3 and 5, respectively. The corresponding p-values are in Tables 4 and 6. Table 7 contains the adjusted CQR estimates, where the adjustment factor is $w = 0.2$. The p-values of Table 4 also apply to the estimates of Table 7. Finally, Table 8 presents the smoothed values of the regressors that are used to construct the predicted values shown in Figures 2 and 3. The coefficient estimates provided in Tables 3, 5, and 7 are at the heart of the price predictions provided in Figures 1 through 3. As discussed in section 2, any differences in the predictions for the same coefficients result from the fact that different values of the

[7] For binary variables, the coefficient is approximately equal to the predicted percentage change in price that results from switching the variable from "off" to "on".

regressors are combined with the coefficients; their mean values in Figure 1, their smoothed values (Table 8) in Figure 2, and a weighted average of their means and smoothed values in Figure 3.

The coefficients of Tables 3 and 7 can be interpreted as approximations of the implicit prices of the characteristics of the single family homes that are being analyzed. As discussed in section 2, the coefficients of Table 5 for the unconditional quantile estimates (UQR) should not be interpreted in this manner. The implicit prices show a significant degree of variation across both the estimation methods and the quantiles for a given estimation method. Figures 4 through 7 visualize these differences for a number of coefficients. Each of the figures provides, for a particular coefficient, the quantile-specific estimates of the traditional conditional quantile regression (CQR), the modified conditional quantile estimate suggested in this study (CQR adjusted), and the unconditional quantile estimates (UQR) derived by Firpo et al. (2007). In addition, each figure contains the OLS estimate as a reference point.

Figures 4 through 7 have in common that the UQR estimates show very large variations across the quantile points. This follows directly from the fact that the coefficients are chosen to map the mean values of the regressors, which are constant across the sample, into the unconditional quantile points of the dependent variable, which vary significantly across quantiles. The UQR method forces all variation of the dependent variable into variations of the coefficients. This can create large surges in the coefficient values at the tail ends of the distribution of the dependent variable. Such surges are particularly pronounced in Figures 4, 5, and 7.

Of main interest are the coefficients of the standard CQR and the adjusted CQR. As discussed in section 2, application of the adjustment factor $(1-w)$ moderates the variation of the adjusted CQR estimates compared to the standard CQR estimates. In practice, this means that the adjusted CQR coefficients are below the standard CQR estimates if the coefficient estimates are positive, and vice versa for negative coefficient values. Figures 4, 5, and 6 illustrate this point. In Figures 4 and 5, the adjusted coefficients are located below the standard ones, in Figure 6, where the coefficients are negative, the adjusted coefficients are above their standard counterparts. In Figure 7, where the CQR coefficients change sign across the quantiles, the line connecting the adjusted CQR coefficients is flatter than the one for the standard CQR.

Apart from the differences across alternative estimation methods, a number of other estimation results are noteworthy. First among them is the fact that the spatial lag variable is statistically highly significant throughout the quantiles and across different estimation methods. This implies that spatial dependence does play a non-negligible role and should not be left out in studies of this type. As for the effect of *age* on the price of a house, a negative sign is expected and returned by all estimation methods. Interestingly, the CQR estimates suggest that age plays much less of a role for expensive homes than for inexpensive ones. This may have to do with the fact that expensive homes tend to be kept up better over time. Very similar to *age* in its effect on price is the variable *winter*. Listing a house in the winter reduces the price but significantly more so for less expensive homes.

The coefficients of the variable *bedroom* are interesting because the literature on real estate valuation contains many conflicting results for this variable, not only with regard to coefficient size but also with respect to sign (Sirmans et al. 2005). The estimates presented in Tables 3 and 7 and illustrated in Figure 7 suggest a possible explanation for the diversity in results. At the lower end of the price distribution, the marginal effect of this variable on price is effectively zero. However, at the upper end of the distribution, the impact is decisively negative. The negative sign often generates confusion in the literature. It is possible to explain this seeming anomaly as follows: carving another bedroom out of a given amount of square footage adds negative value given that high-priced homes already have a sizable number of bedrooms (Table 8); it would reduce the amount of open space available for other purposes, such as entertainment, that are more likely to be highly valued by owners of expensive homes than owners of inexpensive ones.

The various high school zoning variables are largely statistically significant. Some brief explanation will clarify the sign and relative statistical significance of the various zones. The base school zone, that is the zone that is not represented by a variable, is the newest of the county high schools. It is also the one with the fewest minority and low family income students. The three high schools identified as *blkhigh*, *oakhigh*, and *rivhigh* are in direct competition with this new school. All three of these competing high schools have large negative coefficients, which suggests that houses in the zone of the new high school sell at a statistically significant premium.

The last variable *yr2007* identifies the rate of housing price inflation in 2007 relative to 2006. The estimated coefficients presented in Tables 3 and 7 suggest that housing price inflation is below average at the lower end of the price distribution. The impact of inflation is about the same for medium-priced and high-priced homes. On another note, it is interesting to see that prices still appreciated in 2007 given that housing prices were already in decline in other parts of the country during this first year of the bursting of the housing price bubble in the U.S.

5 Summary and Conclusions

The purpose of this paper has been to offer a simple approximate solution to some apparent limitations of conditional quantile regression for the interpretation of hedonic price functions in real estate valuation. These limitations are illustrated by juxtaposing the traditional method of conditional quantile regression, which was introduced by Koenker and Basset (1978), with the method of unconditional quantile regression, which was recently developed by Firpo et al. (2007). Both methods are then compared to the modification of the traditional conditional quantile estimator that is suggested in this study. This modification has the objective to remove the key problem of using conditional quantile regression to derive the implicit valuations of housing characteristics via hedonic price functions: its inability to predict the unconditional quantile values of the dependent variable. The suggested modification of the conditional quantile estimator involves a simple adjustment factor. The adjustment factor lies between zero and one and is particular to a given data set. Fortunately, the appropriate adjustment factor can be derived with relatively little effort.

How the suggested methodology can be applied in practice is illustrated on a data set consisting of transactions data on nearly 5,000 home sales in Tennessee for the years 2006 and 2007. The estimates also show that the unconditional quantile estimator derived by Firpo et al. (2007) is not a useful technique for hedonic price functions applications in real estate valuation. It does not allow to mark particular properties to market or to identify implicit prices for the property's characteristics.

It is apparent that the suggested modification of the conditional quantile estimator is relevant for applications of conditional quantile regression outside of real estate valuation on the basis of hedonic price functions. In particular, the modified estimator is useful for all applications of quantile regression to hedonic price functions because their main purpose is to identify the implicit prices of characteristics. More broadly, the modified estimator should be applied wherever the focus is on identifying differences in the coefficients across quantiles and where these differences form the basis for decision making or intervention at the level of the corresponding individual observation.

References

Anselin, L. (2001). Spatial Econometrics, in *A Companion to Theoretical Econometrics*, B.H. Baltagi ed, Malden, Mass. and Oxford: Blackwell, 310-330.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. New York: Springer.

Billett, M.T., and H. Xue. (2007). The Takeover Deterrent Effect of Open Market Share Repurchases, *Journal of Finance*, 62, 1827-1850.

Bassett, G.W., and H. Chen. (2001). Portfolio Style: Return-based Attribution Using Quantile Regression, *Empirical Economics*, 26, 293-305.

Chernozhukov, V., and C. Hansen. (2008). Instrumental Variable Quantile Regression: A Robust Inference Approach, *Journal of Econometrics*, 142, 379-398.

Fattouh, B., L. Harris, and P. Scaramozzino. (2008). Non-linearity in the Determinants of Capital Structure: Evidence from UK Firms, *Empirical Economics*, 34, 417-438.

Firpo, S., N.M. Fortin, and T. Lemieux. (2007). Unconditional Quantile Regressions. NBER Technical Working Paper 339, Cambridge, M.A.: National Bureau of Economic Research, July.

Horowitz, J.L., and S. Lee. (2007). Nonparametric Instrumental Variables Estimation of a Quantile Regression Model, *Econometrica*, 75, 1191-1208.

Kim, T.-H., and C. Muller. (2004). Two-stage Quantile Regression when the First Stage is Based on Quantile Regression, *Econometrics Journal*, 7, 218-231.

Koenker, R., and K.F. Hallock. (2001). Quantile Regression, *Journal of Economic Perspectives*, 15, 143-56.

Koenker, R., and G. Bassett. (1978). Regression Quantiles, *Econometrica*, 46, 33-50.

Margaritis, D., and M. Psillaki. (2007). Capital Structure and Firm Efficiency, *Journal of Business Finance and Accounting*, 34(9-10), 1447-1469.

Martins, P., and P.T. Pereira. (2004). Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries, *Labour Economics*, 11, 355-371.

McGuinness, S., and J. Bennett. (2007). Overeducation in the Graduate Labour Market: A Quantile Regression Approach. *Economics of Education Review*, 26, 521-531.

Miranda, A. (2008). Planned Fertility and Family Background: A Quantile Regression for Counts Analysis, *Journal of Population Economics*, 21, 67-81.

Nguyen, B.T., J.W. Albrecht, S.B. Vroman, and M.D. Westbrook. (2007). A Quantile Regression Decomposition of Urban-Rural Inequality in Vietnam, *Journal of Development Economics*, 83, 466-490.

Nikolaou, K. (2008). The Behaviour of the Real Exchange Rate: Evidence from Regression Quantiles, *Journal of Banking and Finance*, 32, 664-679.

Pace, R.K., R. Barry, O.W. Gilley, and C.F. Sirmans. (2000). A Method for Spatial-Temporal Forecasting with an Application to Real Estate Prices, *International Journal of Forecasting*, 16, 229-246.

Peng, L., and Y. Huang. (2008). Survival Analysis with Quantile Regression Models, *Journal of the American Statistical Association*, 103, 637-649.

Rangvid, B.S. (2007). School Composition Effects in Denmark: Quantile Regression Evidence from PISA 2000. *Empirical Economics*, 33, 359-388.

Sirmans, G.S., D.A. Macpherson and E.N. Zietz. (2005). The Composition of Hedonic Pricing Models, *Journal of Real Estate Literature*, 13, 3-46.

Taylor, J.W. (2008). Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall, *Journal of Financial Econometrics*, 6, 382-406.

Winkelmann, R. (2006). Reforming Health Care: Evidence from Quantile Regressions for Counts, *Journal of Health Economics*, 25, 131-145.

Yasar, M., and P.C.J. Morrison. (2007). International Linkages and Productivity at the Plant Level: Foreign Direct Investment, Exports, Imports and Licensing, *Journal of International Economics*, 71, 373-388.

Zietz, J., E.N. Zietz, and G.S. Sirmans. (2008). Determinants of House Prices: A Quantile Regression Approach, *Journal of Real Estate Finance and Economics*, 37, 317-333.

TABLE 1. VARIABLE DEFINITIONS

Variable	Definition
lnsp	Log of the sales price; dependent variable
splag	spatial lag variable; average price of houses sold in a 0.3 mile range in the past
sqft	Size of house in square feet, divided by 1,000
lotacre	Lot size in acres
age	Age of house
bedroom	Number of bedrooms
bathf	Number of full bathrooms
deck	1 if deck is present, 0 otherwise
patio	1 if patio is present, 0 otherwise
garage	Garage capacity, in cars
flfinwd	1 if finished wood flooring is present in house, 0 otherwise
exwood	1 if exterior trim is wood, 0 otherwise
exvinyl	1 if exterior trim is vinyl, 0 otherwise
cityw	1 if water source is city water, 0 otherwise
winter	1 if listed in winter season, 0 otherwise
discenter	Distance to city center, in miles
disnash	Distance to downtown Nashville, TN, in miles
blkhigh	1 if school district is Blackman High School, 0 otherwise
eaghigh	1 if school district is Eagleville High School, 0 otherwise
lavhigh	1 if school district is Lavergne High School, 0 otherwise
oakhigh	1 if school district is Oakland High School, 0 otherwise
rivhigh	1 if school district is Riverdale High School, 0 otherwise
smyhigh	1 if school district is Smyrna High School, 0 otherwise
yr2007	1 if transaction year is 2007, 0 otherwise

TABLE 2. BASIC STATISTICS, 4990 OBSERVATIONS

Variable	Mean	Std. Dev.	Min	Max
lnsp	11.9705	0.3364	10.2036	13.7321
splag	1.0684	0.8726	0	6.3500
sqft	1.8564	0.7054	0.6	10.3580
lotacre	0.4670	0.9184	0	36.6800
age	11.7170	11.1119	0	106
bedroom	3.1900	0.5293	0	7
bathf	2.0916	0.4939	1	5
deck	0.4244	0.4943	0	1
patio	0.3830	0.4862	0	1
garage	1.6764	0.7427	0	9
flfinwd	0.4611	0.4985	0	1
exwood	0.0637	0.2443	0	1
exvinyl	0.7399	0.4387	0	1
cityw	0.9567	0.2035	0	1
winter	0.1786	0.3830	0	1
discenter	4.3763	2.9200	0.1638	24.5983
disnash	27.2883	5.3939	14.4049	43.6479
blkhigh	0.1533	0.3603	0	1
eaghigh	0.0032	0.0565	0	1
lavhigh	0.0319	0.1757	0	1
oakhigh	0.1633	0.3697	0	1
rivhigh	0.3321	0.4710	0	1
smyhigh	0.1725	0.3779	0	1
yr2007	0.3251	0.4684	0	1

TABLE 3. CONDITIONAL QUANTILE ESTIMATES, BY QUANTILE, 4990 OBSERVATIONS

	Quantile								
Variables	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.054	0.052	0.048	0.045	0.043	0.041	0.039	0.037	0.032
sqft	0.301	0.330	0.346	0.356	0.366	0.377	0.392	0.398	0.413
lotacre	0.022	0.020	0.032	0.031	0.033	0.031	0.032	0.032	0.042
age	-0.007	-0.006	-0.005	-0.005	-0.004	-0.004	-0.003	-0.003	-0.003
bedroom	0.014	0.000	-0.003	-0.010	-0.013	-0.012	-0.016	-0.016	-0.022
bathf	0.006	0.007	0.007	0.014	0.018	0.017	0.017	0.018	0.033
deck	0.037	0.028	0.021	0.015	0.010	0.006	0.003	0.002	-0.001
patio	0.012	0.011	0.007	0.007	0.007	0.004	0.001	0.004	0.004
garage	0.067	0.061	0.058	0.057	0.055	0.054	0.053	0.054	0.050
flfinwd	0.058	0.047	0.045	0.044	0.042	0.039	0.037	0.038	0.042
exwood	-0.009	-0.011	-0.010	-0.001	-0.001	-0.002	-0.003	0.001	0.008
exvinyl	-0.011	-0.013	-0.016	-0.014	-0.013	-0.016	-0.013	-0.014	-0.020
cityw	-0.006	-0.013	-0.006	-0.010	-0.010	-0.009	-0.019	-0.022	-0.007
winter	-0.027	-0.028	-0.026	-0.023	-0.020	-0.015	-0.016	-0.015	-0.008
discenter	-0.002	-0.002	-0.002	-0.002	-0.001	-0.001	0.000	0.000	-0.001
disnash	-0.003	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	0.000	-0.001
blkhigh	-0.045	-0.044	-0.031	-0.029	-0.026	-0.024	-0.024	-0.018	-0.006
eaghigh	-0.038	-0.070	-0.080	-0.079	-0.044	-0.010	-0.015	-0.024	-0.039
lavhigh	0.011	0.008	0.011	0.015	0.008	0.014	0.012	0.017	0.014
oakhigh	-0.034	-0.038	-0.037	-0.031	-0.028	-0.029	-0.032	-0.031	-0.018
rivhigh	-0.017	-0.026	-0.026	-0.026	-0.024	-0.027	-0.030	-0.033	-0.026
smyhigh	-0.021	-0.003	0.004	0.002	0.005	0.002	0.003	0.005	0.002
yr2007	0.021	0.025	0.028	0.033	0.033	0.032	0.031	0.030	0.033
constant	11.203	11.226	11.223	11.239	11.242	11.237	11.248	11.236	11.230

TABLE 4. P-VALUES FOR CONDITIONAL QUANTILE ESTIMATES OF TABLE 3

	Quantile								
Variables	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sqft	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
lotacre	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
age	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
bedroom	0.164	0.968	0.533	0.015	0.000	0.000	0.000	0.000	0.000
bathf	0.560	0.244	0.235	0.002	0.000	0.000	0.000	0.000	0.000
deck	0.000	0.000	0.000	0.000	0.001	0.060	0.348	0.562	0.813
patio	0.076	0.008	0.137	0.045	0.029	0.143	0.666	0.239	0.328
garage	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
flfinwd	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
exwood	0.531	0.183	0.280	0.840	0.892	0.785	0.694	0.879	0.317
exvinyl	0.156	0.003	0.002	0.000	0.000	0.000	0.000	0.000	0.000
cityw	0.688	0.142	0.534	0.161	0.102	0.133	0.005	0.003	0.423
winter	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.077
discenter	0.154	0.010	0.015	0.002	0.007	0.041	0.667	0.512	0.324
disnash	0.001	0.002	0.030	0.002	0.000	0.004	0.050	0.703	0.290
blkhigh	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.392
eaghigh	0.396	0.022	0.015	0.002	0.038	0.622	0.507	0.347	0.136
lavhigh	0.573	0.503	0.387	0.127	0.329	0.077	0.198	0.077	0.198
oakhigh	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
rivhigh	0.099	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
smyhigh	0.088	0.674	0.646	0.772	0.286	0.638	0.620	0.386	0.801
yr2007	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 5. UNCONDITIONAL QUANTILE ESTIMATES, BY QUANTILE, 4990 OBSERVATIONS

	Quantile								
Variables	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.029	0.035	0.029	0.030	0.031	0.036	0.045	0.061	0.069
sqft	-0.005	0.066	0.147	0.235	0.388	0.453	0.477	0.523	0.567
lotacre	0.009	0.007	0.005	0.006	0.003	0.008	0.015	0.027	0.057
age	-0.005	-0.003	-0.002	-0.003	-0.004	-0.005	-0.005	-0.007	-0.005
bedroom	0.045	0.008	-0.009	-0.024	-0.032	-0.026	-0.003	0.037	0.050
bathf	0.109	0.043	0.004	-0.025	-0.066	-0.088	-0.096	-0.083	0.049
deck	0.070	0.054	0.049	0.049	0.045	0.045	0.018	-0.027	-0.100
patio	0.028	0.023	0.008	0.005	-0.005	-0.008	-0.007	0.000	0.001
garage	0.122	0.118	0.116	0.109	0.093	0.043	-0.005	-0.031	-0.029
flfinwd	0.017	0.035	0.042	0.061	0.098	0.131	0.152	0.157	0.077
exwood	-0.003	0.021	0.021	0.015	-0.005	-0.009	-0.027	-0.058	-0.035
exvinyl	0.003	0.000	0.001	-0.006	-0.009	-0.021	-0.027	-0.066	-0.072
cityw	-0.012	-0.018	0.002	0.002	-0.024	-0.036	-0.034	-0.006	-0.027
winter	-0.023	-0.019	-0.019	-0.016	-0.007	0.001	-0.002	-0.020	-0.014
discenter	-0.002	-0.003	-0.005	-0.005	-0.002	0.000	0.004	0.001	0.001
disnash	-0.003	-0.003	-0.003	-0.003	0.000	0.000	0.002	0.002	0.000
blkhigh	0.007	0.001	0.047	0.037	0.023	-0.043	-0.079	-0.114	-0.104
eaghigh	0.005	-0.041	-0.061	-0.054	-0.087	-0.074	0.007	0.108	0.021
lavhigh	-0.029	-0.012	0.038	0.064	0.161	0.137	0.056	0.016	-0.168
oakhigh	-0.074	-0.088	-0.057	-0.057	-0.064	-0.060	-0.044	-0.028	0.117
rivhigh	0.035	0.025	0.029	-0.003	-0.042	-0.044	-0.058	-0.092	-0.134
smyhigh	0.034	0.005	0.028	0.018	0.025	-0.013	-0.043	-0.102	-0.095
yr2007	0.025	0.046	0.046	0.042	0.031	0.024	0.019	0.031	0.011
constant	11.117	11.342	11.347	11.370	11.260	11.362	11.394	11.366	11.229

TABLE 6. P-VALUES FOR UNCONDITIONAL QUANTILE ESTIMATES OF TABLE 5

	Quantile								
Variables	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sqft	0.633	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
lotacre	0.038	0.031	0.165	0.116	0.522	0.102	0.002	0.000	0.000
age	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
bedroom	0.000	0.322	0.255	0.005	0.002	0.014	0.786	0.007	0.009
bathf	0.000	0.000	0.639	0.009	0.000	0.000	0.000	0.000	0.024
deck	0.000	0.000	0.000	0.000	0.000	0.000	0.089	0.043	0.000
patio	0.004	0.003	0.288	0.575	0.635	0.437	0.534	0.980	0.967
garage	0.000	0.000	0.000	0.000	0.000	0.000	0.459	0.000	0.014
flfinwd	0.064	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
exwood	0.859	0.159	0.150	0.343	0.787	0.629	0.174	0.026	0.327
exvinyl	0.760	0.978	0.860	0.483	0.372	0.054	0.015	0.000	0.000
cityw	0.552	0.242	0.898	0.907	0.239	0.077	0.111	0.817	0.470
winter	0.029	0.017	0.015	0.070	0.506	0.960	0.841	0.168	0.475
discenter	0.261	0.003	0.000	0.000	0.305	0.895	0.013	0.762	0.598
disnash	0.027	0.000	0.000	0.008	0.920	0.802	0.204	0.250	0.870
blkhigh	0.654	0.906	0.000	0.003	0.131	0.005	0.000	0.000	0.000
eaghigh	0.948	0.463	0.264	0.368	0.231	0.315	0.927	0.268	0.873
lavhigh	0.270	0.564	0.053	0.003	0.000	0.000	0.043	0.644	0.001
oakhigh	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.162	0.000
rivhigh	0.011	0.022	0.006	0.807	0.003	0.002	0.000	0.000	0.000
smyhigh	0.042	0.711	0.027	0.184	0.133	0.441	0.015	0.000	0.003
yr2007	0.008	0.000	0.000	0.000	0.001	0.016	0.069	0.017	0.544
constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 7. ADJUSTED CONDITIONAL QUANTILE ESTIMATES, BY QUANTILE, 4990 OBSERVATIONS

	Quantile								
Variables	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.043	0.042	0.038	0.036	0.035	0.033	0.032	0.030	0.026
sqft	0.241	0.264	0.277	0.285	0.293	0.301	0.313	0.318	0.331
lotacre	0.018	0.016	0.025	0.025	0.026	0.025	0.026	0.026	0.033
age	-0.006	-0.005	-0.004	-0.004	-0.003	-0.003	-0.003	-0.003	-0.002
bedroom	0.011	0.000	-0.003	-0.008	-0.010	-0.009	-0.013	-0.012	-0.018
bathf	0.004	0.005	0.006	0.011	0.014	0.014	0.014	0.015	0.027
deck	0.030	0.022	0.017	0.012	0.008	0.004	0.002	0.002	-0.001
patio	0.010	0.009	0.006	0.006	0.005	0.004	0.001	0.003	0.003
garage	0.053	0.049	0.046	0.045	0.044	0.044	0.042	0.043	0.040
flfinwd	0.046	0.037	0.036	0.035	0.034	0.031	0.029	0.031	0.034
exwood	-0.007	-0.009	-0.008	-0.001	-0.001	-0.001	-0.002	0.001	0.007
exvinyl	-0.009	-0.011	-0.013	-0.012	-0.011	-0.013	-0.011	-0.012	-0.016
cityw	-0.005	-0.010	-0.005	-0.008	-0.008	-0.007	-0.015	-0.017	-0.005
winter	-0.027	-0.028	-0.026	-0.023	-0.020	-0.015	-0.016	-0.015	-0.008
discenter	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000	0.000	-0.001
disnash	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000	0.000
blkhigh	-0.036	-0.035	-0.025	-0.023	-0.020	-0.019	-0.020	-0.014	-0.004
eaghigh	-0.030	-0.056	-0.064	-0.064	-0.035	-0.008	-0.012	-0.019	-0.031
lavhigh	0.009	0.006	0.009	0.012	0.006	0.011	0.009	0.014	0.011
oakhigh	-0.027	-0.031	-0.030	-0.024	-0.023	-0.023	-0.026	-0.025	-0.014
rivhigh	-0.014	-0.021	-0.021	-0.021	-0.019	-0.021	-0.024	-0.026	-0.021
smyhigh	-0.017	-0.002	0.003	0.001	0.004	0.002	0.002	0.004	0.001
yr2007	0.021	0.025	0.028	0.033	0.033	0.032	0.031	0.030	0.033

TABLE 8. SMOOTHED VALUES OF REGRESSOR VARIABLES, 4990 OBSERVATIONS

Variables	Quantile 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
splag	0.7485	0.8333	0.8900	0.9388	1.0241	1.1317	1.2078	1.3040	1.4587
sqft	1.2840	1.3715	1.4716	1.5809	1.6922	1.8814	2.0754	2.3387	2.7286
lotacre	0.4091	0.4148	0.4188	0.4263	0.4299	0.4544	0.4769	0.5189	0.6041
age	15.3854	14.2885	13.4583	12.7255	11.8148	10.7746	9.8589	8.7696	7.3465
bedroom	2.9026	2.9636	3.0139	3.0618	3.1264	3.2085	3.2897	3.3998	3.5735
bathf	1.7965	1.8562	1.9059	1.9536	2.0187	2.1025	2.1866	2.3026	2.4898
deck	0.4114	0.4144	0.4168	0.4191	0.4221	0.4258	0.4295	0.4343	0.4415
patio	0.3562	0.3622	0.3671	0.3717	0.3779	0.3856	0.3931	0.4030	0.4182
garage	1.2898	1.3610	1.4215	1.4806	1.5627	1.6710	1.7824	1.9407	2.2073
flfinwd	0.1875	0.2388	0.2869	0.3366	0.4085	0.5031	0.5941	0.7041	0.8323
exwood	0.0927	0.0833	0.0763	0.0703	0.0630	0.0548	0.0480	0.0402	0.0306
exvinyl	0.7526	0.7498	0.7474	0.7452	0.7423	0.7387	0.7351	0.7304	0.7231
cityw	0.9556	0.9559	0.9561	0.9563	0.9565	0.9568	0.9572	0.9575	0.9581
winter	0.1963	0.1919	0.1884	0.1852	0.1810	0.1759	0.1711	0.1649	0.1559
discenter	4.3400	4.3381	4.3090	4.3389	4.3462	4.3687	4.3573	4.3953	4.5209
disnash	27.8618	27.6148	27.3996	27.2874	27.0870	27.0868	26.9634	26.9636	27.1672
blkhigh	0.1505	0.1511	0.1517	0.1521	0.1528	0.1536	0.1544	0.1554	0.1570
eaghigh	0.0048	0.0042	0.0038	0.0035	0.0031	0.0027	0.0023	0.0019	0.0014
lavhigh	0.0210	0.0228	0.0244	0.0259	0.0281	0.0311	0.0342	0.0388	0.0470
oakhigh	0.1567	0.1582	0.1594	0.1605	0.1620	0.1640	0.1658	0.1683	0.1720
rivhigh	0.3605	0.3537	0.3482	0.3432	0.3365	0.3283	0.3205	0.3104	0.2954
smyhigh	0.2068	0.1978	0.1907	0.1843	0.1760	0.1662	0.1571	0.1458	0.1299
yr2007	0.2768	0.2871	0.2957	0.3038	0.3148	0.3287	0.3424	0.3608	0.3895

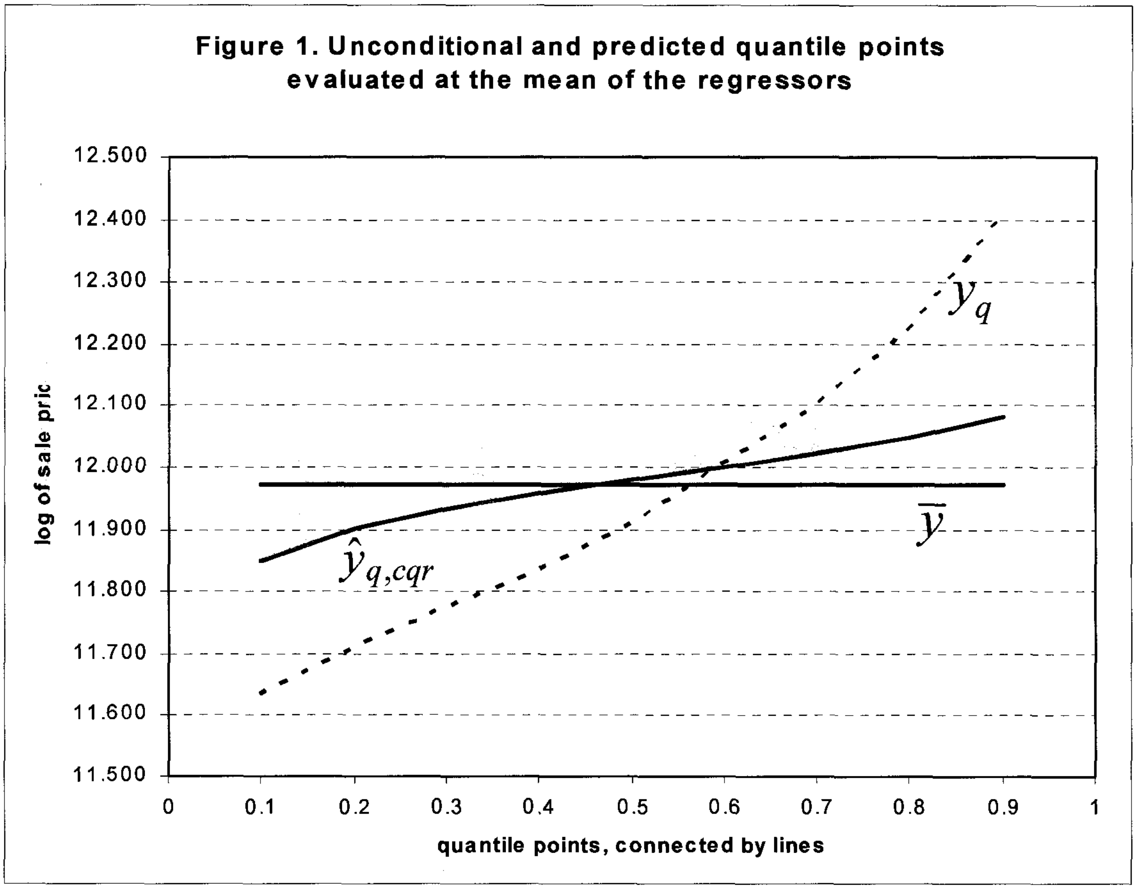


Figure 1. Unconditional and predicted quantile points
evaluated at the mean of the regressors
12.500
12.400
12.300
12.200
12.100
12.000
11.900
11.800
11.700
11.600
11.500
log of sale pric
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
quantile points, connected by lines
y_q
$\bar{y}$
$\hat{y}_{q,cqr}$

Figure 2. Unconditional and predicted quantile points at smoothed values of regressors

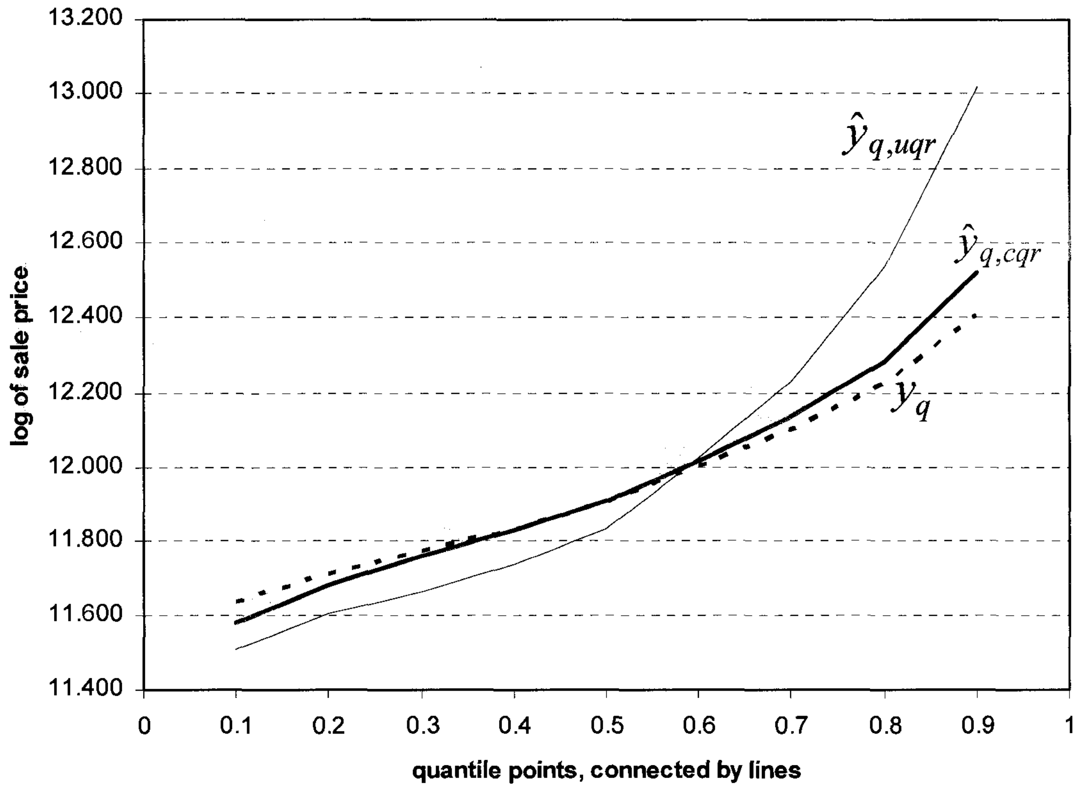


Figure 3. Unconditional and predicted quantile points evaluated at weighted average of mean and smoothed values of regressors

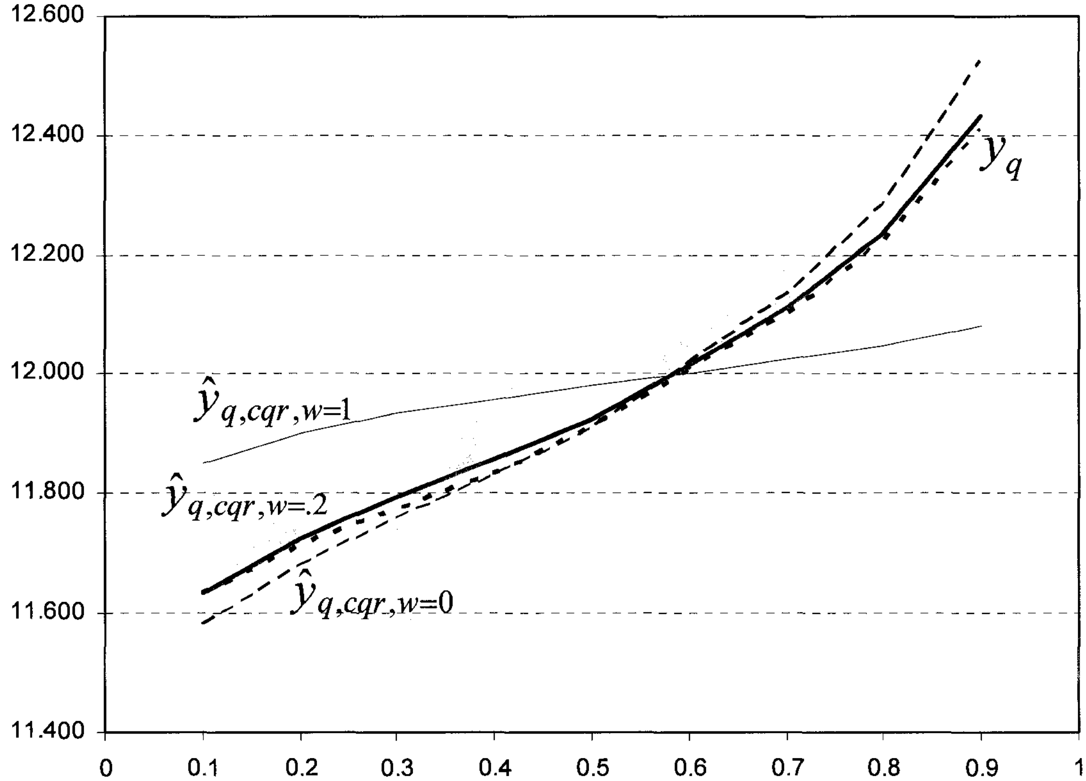


Figure 4. Alternative Coefficients of Variable *sqft*

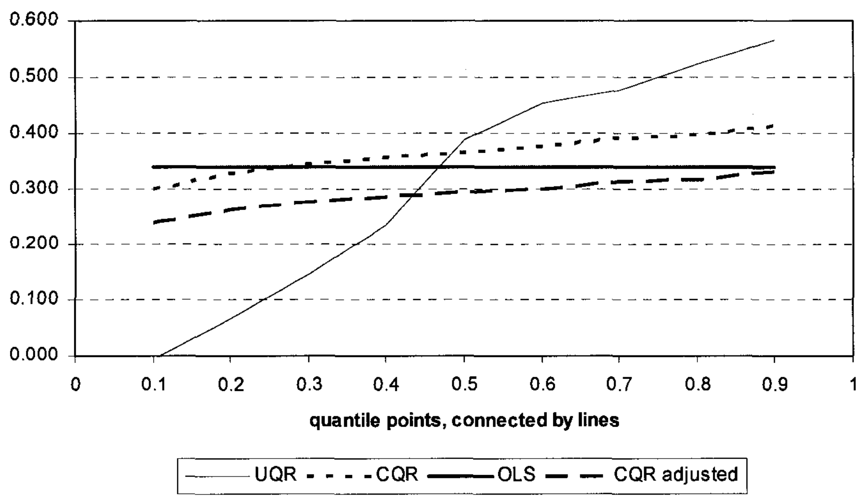
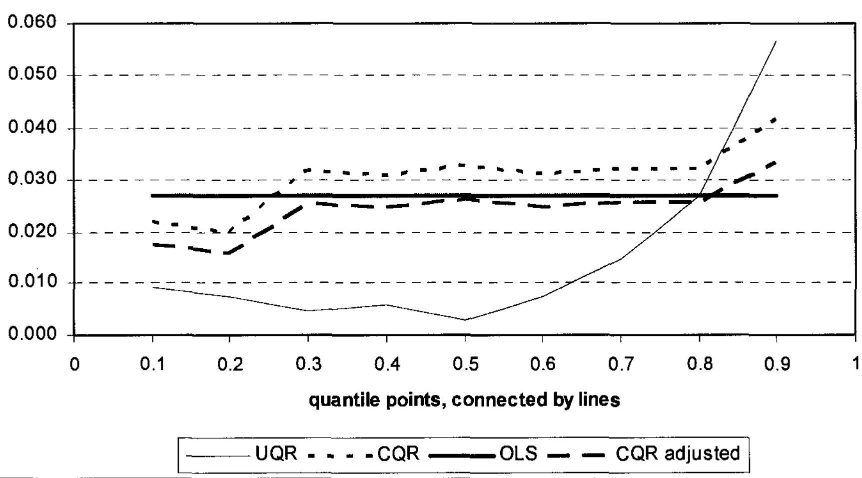


Figure 5. Alternative Coefficients of Variable *lotacre*



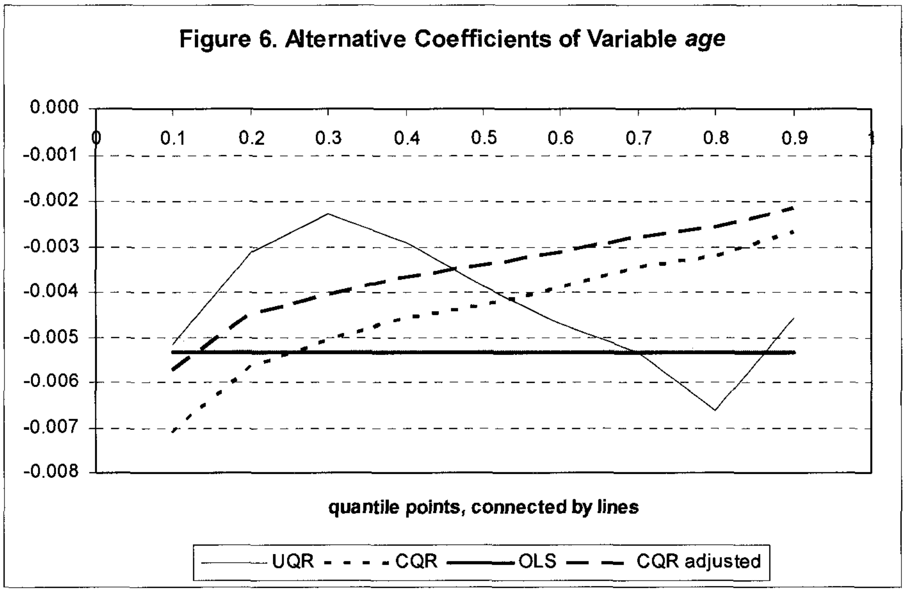


Figure 6. Alternative Coefficients of Variable *age*
0.000
-0.001
-0.002
-0.003
-0.004
-0.005
-0.006
-0.007
-0.008
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
quantile points, connected by lines
UQR
CQR
OLS
CQR adjusted

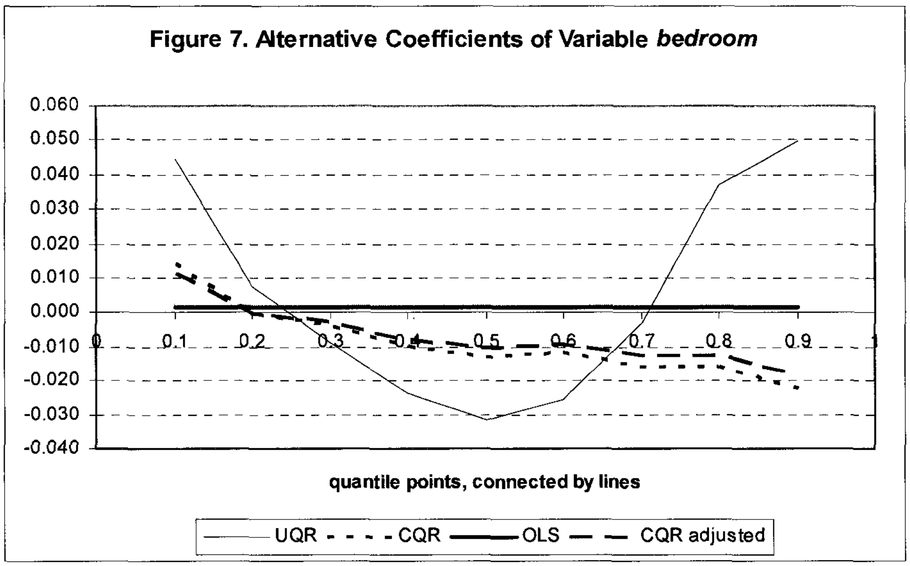


Figure 7. Alternative Coefficients of Variable *bedroom*
0.060
0.050
0.040
0.030
0.020
0.010
0.000
-0.010
-0.020
-0.030
-0.040
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
quantile points, connected by lines
UQR
CQR
OLS
CQR adjusted

TYPE OF FLOORING AS A DETERMINANT OF THE SALE PRICE OF SINGLE-FAMILY HOMES: A QUANTILE REGRESSION APPROACH

Abstract

Earlier hedonic studies of housing markets have established the fact that flooring type is an important determinant of the selling price of a single-family home, but have provided little detail. This study focuses on the impact of a number of specific flooring types on price over a time period of five years. The study departs from the traditional ordinary least squares methodology of estimating hedonic price functions by employing a conditional quantile estimator. This methodology allows the implicit prices of different flooring types to differ across the distribution of home prices. Spatial effects are accounted for through the use of a spatial lag model. The empirical work is based on Multiple Listing Service data of single-family homes sold in Tennessee County for the years 2003 to 2007.

1 Introduction

The rich literature on hedonic price models in real estate valuation has provided a large number of estimates of the implicit price of house characteristics, such as square footage and age (Sirmans, Macpherson and Zietz, 2005). However, there is very little detailed analysis of the implicit prices of different types of flooring. The purpose of this paper is to fill this void in the literature. In particular, using traditional and modified conditional quantile regression methods in a hedonic price framework, this paper attempts to examine the impact of eight major flooring types on the selling prices of single-family homes in a Tennessee county for the years 2003 to 2007.

The motivation of this study is twofold. First, flooring type is undoubtedly an important determinant of house price. Flooring type defines the style of a house and the atmosphere of a room. Builders, buyers, appraisers, and home owners take a careful look at the flooring type when they evaluate a house. It is, therefore, surprising that existing studies have not fully explored the extent to which flooring type will affect selling price. The objective of this paper is to explore the importance of flooring in determining house price in some detail.

Second, a key component of the approach is to allow the implicit prices of major flooring types to vary along the distribution of house price. This is potentially important because there is reason to believe that the owners/buyers of inexpensive homes may value the flooring type or other home characteristics differently from the owners/buyers of expensive ones (Malpezzi 2003). For instance, tile flooring could significantly raise the price of inexpensive homes, while tile flooring may add little to the value of higher-priced homes. This possibility is excluded on a priori grounds in previous studies that included variables on flooring. Yet, it may be important information for the decisions of home owners, builders, and appraisers alike. Zietz et al. (2008) have shown how this information may be extracted from the data through the use of conditional quantile regression (CQR), a methodology originally introduced into the literature by Koenker and Bassett (1978).[1] However, as detailed in Zhou and Zietz (2008), the conditional quantile methodology of Koenker and Bassett (1978) will need to be modified to fully address the needs of real estate valuation with hedonic price functions. This modified estimator will be used and compared to the traditional conditional quantile results and least squares regression.

This study addresses spatial dependence through a spatial lag. The spatial lag is constructed as a weighted average of the selling prices of neighborhood homes that sold before the house in question. This idea was introduced into the literature by Pace et al. (2000). It avoids the endogeneity problem that is typically associated with a spatial lag model and, therefore, makes estimation significantly easier, which is an important consideration in the context of quantile regression.

The findings of this paper should be of interest to home buyers, developers, appraisers, and real estate agents alike. The results show that there are some large differences in the implicit prices of different combinations of flooring. For example, least squares estimates reveal that a home with a combination of carpet, wood, and tile sells for a premium of almost 10% over a house with the most frequently observed flooring combination of carpet and vinyl. The estimates also show that, for certain floor

[1] A convenient recent summary of this methodology is provided by Koenker and Hallock (2001).

combinations, there are perceptible differences in implicit valuations across the quantiles of the selling price. For example, the combination of wood and tile adds almost 11 percent to the value of low-priced homes at the 0.1 quantile, while it the percentage impact is only half as large at the 0.9 percent quantile.

The paper is structured as follows. Section 2 illustrates the methodology used for the estimation of the hedonic price model. The technique for addressing spatial-temporal autocorrelation is also discussed in this section. Section 3 introduces the data and provides descriptive statistics. Section 4 discusses the empirical estimation results. The last section concludes.

2 Methodology

2.1 Ordinary Least Squares

Over the past three decades, least squares regressions (OLS) on hedonic price functions have been used extensively in the housing market literature to identify the effect of housing characteristics on the value of a house. The regression coefficients of these models are interpreted as the implicit prices of the housing attributes (Freeman 1979).

The theory of hedonic models requires that a full set of the potential determinants of housing price should be included as independent variables. Accordingly, aside from the nine variables that identify different types of flooring (F), which are the focus of this study, two other categories of independent variables (X) are included in the models. The first category includes the structural characteristics of the property, such as house size and lot size. The second category contains an indicator of prices in the neighborhood, some distance variables, and dichotomous variables such as the year the property was sold and the season it was initially listed. The semi-logarithmic functional form is preferred in the hedonic literature due to its easy interpretation and good fit (Halvorsen and Palmquist 1980). It is also used in this study. Hence, the mathematical expression for the equation to estimate is given as

$$\ln P = \alpha + \beta_F F + \beta_x X + \varepsilon \,, \ \varepsilon \sim NID(0, \sigma^2) \,, \tag{1}$$

where the error term ε is assumed to be normally distributed with mean zero and constant variance (σ^2). The coefficients β_F and β_X capture the corresponding implicit prices of the covariates.

Heteroskedasticity and multicollinearity are two frequent concerns in OLS hedonic price estimation and need to be examined. Heteroskedasticity occurs when the variance of the residuals from the regression is not constant. Though presence of heteroskedasticity does not necessarily bias the coefficients, it does give rise to inefficient estimation. The Breusch-Pagan/Cook-Weisberg test is used to check for the presence of heteroskedasticity. For the given data set, a very small p-value (<0.0000) suggests the presence of heteroskedasticity. Therefore, robust standard errors are used for OLS estimation throughout.

Multicollinearity arises when there is a linear relationship among two or more predictors. It can be detected by large values of the variance inflation factors (VIF). As a rule of thumb, a variable whose VIF value is greater than 10 generally indicates severe multicollinearity. In this model, we calculate VIF for all variables and have a mean VIF value of 1.90, which suggests that multicollinearity is not a major problem.[2]

2.2 *Conditional Quantile Regression*

The primary reason for the extensive application of OLS regression for hedonic price models is its straightforward interpretation. However, the OLS estimator is rather sensitive to even a modest number of outliers, and thus, can provide misleading conclusions. Quantile regression, first introduced by Koenker and Bassett (1978), generates more robust estimators when there are outliers.[3] Robustness is an important estimator property when the data set consists of a large set of observations and when these observations are not very reliable. This is routinely the case for the data sets used in real estate valuation. They draw heavily on data from multiple listing services (MLS), which have only a moderate degree of reliability even after apparent data errors are eliminated based on plausibility checks.

As discussed in some detail in Zietz et al. (2008), quantile regression is of interest for the estimation of hedonic price functions not only for its robustness property. It also allows the researcher to identify different implicit prices for different quantiles of the price distribution. This makes it possible, for example, to check whether the implicit price of an additional square footage of living space is valued the same for lower priced homes as for higher priced homes. Such information is of interest to all parties that rely on

[2] The maximum and minimum VIF values are 6.54 and 1.01 respectively.
[3] Quantile regression refers to conditional quantile regression in this paper.

accurate valuations for their decision making, such as buyers, sellers, builders, real estate brokers or others.

In quantile regression, the coefficients of the hedonic price model are estimated separately for each quantile point of the distribution of the dependent variable that one wishes to analyze. Regardless of the particular quantile point that one is interested in, all sample observations are always employed for estimation purposes. No observations are omitted. Different estimators come about because the same observations are weighted differently by quantile regressions that are focused on different quantile points. For each quantile, a new set of coefficients emerges. Estimating equation 1 for the q^{th} quantile point involves selecting the coefficient set $\{\alpha, \beta_F, \beta_x\}$ to minimize the weighted sum of the absolute deviations,

$$\min_{\{\alpha,\beta_F,\beta_X\}} \sum_i |\varepsilon_i| h_i ,$$

where ε_i denotes the residual at observation *i*,

$$\varepsilon_i = \ln P_i - \alpha - \beta_F F_i - \beta_x X_i ,$$

and where h_i is a weight defined as

$$h_i = 2q$$

if ε_i is strictly positive or as

$$h_i = 2 - 2q$$

if ε_i is negative or zero. The coefficients set $\{\alpha, \beta_F, \beta_x\}$ will typically change depending on the choice of quantile q $(0 < q < 1)$. The extent to which the coefficients change is determined by the data.

The coefficients from conditional quantile regression are interpreted as the responses of the *conditional* quantile values of the dependent variable to a unit change in the mean value of the independent variable. For OLS, the mean value of the sample observations of the dependent variable is equal to the predicted value of the dependent variable ($\hat{y}$), if the latter is obtained as the sum over all j regressors of the products of the estimated coefficients ($\hat{\beta}_{j,OLS}$) and the mean values of the corresponding regressors ($\bar{X}_j$),

$$\bar{y} = \hat{y} = \sum_j \hat{\beta}_{j,OLS} \bar{X}_j .$$

Similarly, the q^{th} conditional quantile value of the dependent variable can be predicted for a given set of covariates from the estimated coefficients ($\hat{\beta}_{j,CQR}$) and the mean values of the corresponding regressors ($\bar{X}_j$),

$$\hat{y}_q = \sum_j \hat{\beta}_{j,CQR} \bar{X}_j .$$

However, the conditional quantile value $\hat{y}_q$ is not necessarily equal to the unconditional or observed quantile value of y. This conclusion does not change if the regressors are evaluated at those values that correspond to the chosen quantile of the dependent variable rather than at their mean values. This point is illustrated in Figure 1. The green thick line in Figure 1 plots the observed unconditional quantile values ($0 < q < 1$) of the selling prices. The blue line represents the mean of the sales price distribution. It is also the predicted value from the OLS regression if the regressors are evaluated at their mean values. The thin red line represents the predicted values from the conditional quantile regression with the regressors evaluated at their mean values. It is apparent that the predicted values move away from the actual unconditional quantile values as house prices move away from the median ($q = 0.5$); that is, the unconditional quantile prices tend be overestimated at lower quantiles and overestimated at higher quantiles. By contrast, the broken line provides a much closer fit of the unconditional quantile values, although there remains a systematic bias. The broken line represents the conditional quantile predictions when the regressors are evaluated at those values that are typical for the given quantiles of the selling price. Regressor values that are typical for the q^{th} quantile point are derived as predicted values of inverse regressions, where the regressor is the dependent variable and the selling price the independent variable.[4]

2.3 Adjusted Conditional Quantile Regression

Based upon the two conditional quantile regressions depicted in Figure 1, it is natural to expect that a better approximation of the unconditional quantile points can be achieved by averaging the two quantile predictions. That way, the observable over- and under-predictions will potentially counter each other. As suggested in Zhou and Zietz (2008), better predictions of the unconditional quantile points can be achieved by multiplying the standard conditional quantile estimates with a weighted average ($\widetilde{X}$) of both the mean and the smoothed values of the regressors that are associated with a particular quantile point,

$$\widetilde{X} = w\overline{X} + (1 - w)\hat{X} \ ,$$

[4] Lowess regression is used for continuous regressors and logit regressions for binary regressors. The predictions of these regressions are the "smoothed" regressor values at quantile q. Details are provided in Zhou and Zietz (2008).

where w ($0 \leq w \leq 1$) is the weight to be determined by the data. Setting $w = 1$, the predictions would be identical to those of the traditional conditional quantile regression evaluated at the mean values of the regressors and identified as the thin red line in Figure 1. By contrast, setting $w = 0$, the predictions would be those of the broken line in Figure 1. Apparently, a value of $0 < w < 1$ is likely to provide better predictions of the unconditional quantile points.

Once one accepts the above logic, the q^{th} *unconditional* quantile can be approximated by the equation

$$\sum_j \hat{\beta}_{j,CQR} \tilde{X}_j = \sum_j \hat{\beta}_{j,CQR} \left(w\bar{X}_j + (1-w)\hat{X}_j \right).$$

The equation implies that the marginal effect of a small change in $\hat{X}_j$ on the unconditional quantile point is given as

$$\hat{\beta}_{j,CQR}(1-w).$$

This result assumes that a small change in $\hat{X}_j$ does not affect the mean value of X_j.[5] If a change in $\hat{X}_j$ automatically implies a change in $\bar{X}_j$, as for example for time trend variables, seasonal dummy variables, or the regression constant, no adjustment is needed. To sum up, approximately correct predictions of a given unconditional quantile point can be obtained if one adjusts the traditional conditional quantile estimator by the factor (1 – *w*) for most variables other than the constant term and variables that identify the time of the sale.

[5] For variables that are not house specific, such as the regression constant or a binary variable that captures the rise in general inflation, the adjustment factor needs to be zero because the mean value of these variables would change if one assumed the variable changed for a particular quantile.

2.4 Spatial Autocorrelation and Spatial-Temporal Technique

The importance of spatial autocorrelation in hedonic housing price function analysis has been widely discussed (Dubin 1988, 1992; Can 1990, 1992; Bowen, Mikelbank, and Prestegaard 2001; Clapp and Tirtiroglu 1994; Holly, Pesaran and Yamagata 2007). At its simplest, spatial autocorrelation refers to the connection between houses, subdivisions, census tracts, and regions due to a similar location in space. With more distance between objects, spatial autocorrelation should decline. House prices are likely to be spatially correlated because neighborhoods tend to have similar houses and amenities. Spatial autocorrelation can generate erroneous coefficient estimates and statistical tests in OLS regressions (Dubin 1992). Not all studies appear to have serious spatial autocorrelation issues. Basu and Thibodeau (1998), for example, report effectively no spatial autocorrelation.

Incorporating spatial autocorrelation can complicate the estimation of hedonic price functions. This applies, in particular, if estimates rely on conditional quantile regression. Zietz et al. (2008) employ a two-stage estimator to remove the endogeneity that is introduced by a spatial lag representation. However, the mechanics of that estimator are relatively involved. This study uses an alternative approach to estimate a spatial lag model: it makes use of the spatial-temporal technique introduced by Pace et al. (2000). The basic idea of this technique is to generate for each house a "neighbor candidate pool" that is composed of all houses that are sold before the given house and are not too far away in terms of distance. To guarantee that the "neighbor candidate pool" is large enough for the houses sold in the first year of the sample, data from a previous year (2002) are added for the purpose of neighbor selection. The selected neighbors for each observation in the sample are then re-evaluated based on the distance to the given house. Neighbors located beyond 0.5 mile are disregarded. A proximity weight can be calculated for each neighbor from the ratio of the inverse of its distance and the sum of the inverse distances to all neighbors. A weighted neighborhood price for a given house is generated from the proximity weights and the selling prices of the neighbors. This spatial lag variable (*slag_lnsp*) entering the regression. Since only prices of neighboring houses enter the spatial lag variable that sold before the given house, the spatial lag variable can be considered exogenous. As a consequence, OLS or conditional quantile regression can be applied without having to account for endogeneity bias.

3 Data

The analysis is based on Multiple Listing Service (MLS) data collected by the Middle Tennessee Regional Multiple Listing Service (MTRMLS). The data contain information on single-family house transactions in Rutherford County, Tennessee, over the years 2003 through 2007. Table 1 presents the variables and their definitions. Table 2 provides descriptive statistics for each variable for the full sample of 6,251 observations.

Eight major types of flooring are identified, carpet, wood, tile, vinyl, marble, slate, parquet, and laminate. All other flooring types are categorized as "other." Each flooring type is represented by a binary variable, with one representing the presence of the flooring type. Typically, homes have more than one flooring type. Measurement of flooring type by area covered would be desirable, but is not available.

Table 3 summarizes the percentage of the presence of each flooring type in the full sample and by quantile of sales price. Almost all houses across the quantiles of sales price have carpet flooring present in the house, with the percentage ranging from 91.56% to 98.05%. Vinyl is the second most popular flooring type, with 71.24% for the sample. The percentages for vinyl flooring decline dramatically as one moves toward the more expensive homes, from 85.94% for houses at the 0.1 quantile to 24.2% for houses at the 0.9 quantile. For the whole sample, 42.68% houses have wood flooring. For lower-priced houses, however, less than 30% houses have wood flooring. The percentage is much higher in higher-priced houses. A similar pattern can be found for tile flooring: as house prices increase, a larger percentage of houses have tile flooring. Marble is an expensive flooring material and not very common in the sample. Less than 1% of the houses in the sample have marble flooring. This applies across all quantiles, except the most expensive homes, of which 3.21% have marble flooring. Slate and laminate are rarely used flooring materials. Parquet is a geometric mosiac of wood pieces used for decorative effect. It is somewhat more prevalent in medium-priced homes.

Several structural characteristics are used as covariates. Among the eleven structural characteristics variables are such common variables as square footage, lot size, the number of bedrooms, and the number of full and half bathrooms. The age of the house enters with a linear and squared term to allow for a nonlinear effect of age on sale price.

To account for locational attributes of the houses, all homes are geo-coded. For each house, two distance variables are calculated. The first distance variable is the distance (in miles) to the center of the city where the house is located. The second distance variable is the distance to downtown Nashville, TN. It is expected that a close distance to Nashville should serve to increase house sales price. Both distance variables are calculated using the Grand Circle distance formula:

$$\text{distance} = r \times \arccos[\sin(lat1/\varphi) \times \sin(lat2/\varphi) + \cos(lat1/\varphi) \times \cos(lat2/\varphi) \times \cos(lon2/\varphi - lon1/\varphi)],$$

where r is the radius of the earth (3,963 miles). The scalar $\varphi = 180/\pi$ is used to convert the decimal coordinates to radians. The variables $lat1$ and $lon1$ represent the latitude and longitude of one location and $lat2$ and $lon2$ represent the latitude and longitude of the other location.

Year dummy variables are constructed for the year in which the house is sold. They control for general inflation over the study period from 2003-2007. The base year is set to 2003. Therefore, four dummy variables are incorporated in the regression. Finally, a binary variable for the winter season (*winter*) controls for the impact of the listing season on sales price. Specifically, if the house is listed in November, December, or January, the variable *winter* equals unity.

4 Results

4.1 Standard Least Squares and Conditional Quantile Results

The coefficient estimates from OLS and standard conditional quantile regression are provided in Table 4; the corresponding p-values are given in Table 5.[6]

[6] The p-values for OLS are based on heteroskedasticity robust standard errors.

Tables 4 and 5 reveal that carpet flooring does not play a significant role in determining the sales prices of homes, except for some lower quantiles. This can be explained by the fact that carpet is present in almost all houses across the quantiles. By contrast, vinyl, wood, and tile flooring are statistically significant for all quantiles. Vinyl flooring is the only type of flooring with a consistently negative impact on house prices for all quantiles. The coefficients vary surprisingly little across quantiles, which means that vinyl flooring is considered inferior across all price ranges. The presence of vinyl can be expected to lower the price of a house on average by two percent. Of all flooring types, wood has the strongest positive impact on house prices. At the lower quantiles, where wood is not used very frequently (Table 3), wood flooring is likely to raise price from four to six percent. Its marginal impact decreases for the upper quantiles where wood is already heavily in use. The conclusions for tile flooring are similar to those for wood flooring, although tile flooring does not capture quite the same price premium that wood does. Parquet flooring adds modest value only for the lowest three quantiles of houses. Marble has a statistically significant positive impact only for a few select quantiles; the predicted percentage changes are similar to those for tile. Slate, laminate, and other flooring are not statistically significant across the quantiles.

The quantile estimates of Table 4 show perceptible variation in the non-flooring coefficients across quantiles, a result similar to those reported in Zietz et al. (2008) for a much smaller data set. For example, an increase in house square footage, lot size, or full bathrooms raises price for all houses, but the increase is far greater at the upper end of the house price distribution.

A number of housing characteristics have a statistically significant negative impact on price as well as noticeable coefficient differences across quantiles. For example, an additional bedroom lowers the price at the 0.9 quantile by almost 2%, whereas it has effectively no impact at the lower end of the price distribution. This result suggests that buyers of expensive homes value large rooms and open space more than buyers of inexpensive homes. Vinyl on the exterior of the house also subtracts far more from the value of expensive homes than of inexpensive homes. By contrast, the negative effect of age on price is more pronounced for inexpensive than expensive homes. One may presume that owners of expensive homes have more of an incentive to invest in upkeep and maintenance. At the 0.2 quantile, an additional year lowers the price by almost half a percent, while the price at the 0.9 quantile point is lowered by only 0.3%.[7]

The spatial lag variable *slag_lnsp* is statistically not significant at any quantile, although the coefficients are far larger at the lower end of the price distribution. The results suggest that spatial correlation is not very important for the data set at hand. Being close to the city center appears to matter far more for homes at the lower price range, while a larger distance to Nashville is estimated to be equally unfavorable to all homes, except those at the very bottom quantile. Homes listed originally during the winter months sell for less. This effect is quite perceptible for lower priced homes. The negative effect of the winter listing months on price declines toward the upper quantiles and is effectively absent for the top quantile of homes.

[7] The predicted price change from one additional year for the 0.2 quantile is -0.00484, which is calculated as (-0.0033) + 2(-0.000067)(11.5), where 0.0033 is the coefficient of the variable *age*, 0.000067 the coefficient of the variable *age2* divided by 100 (see Table 1), and 11.5 the sample mean of the variable *age*.

The impact of general inflation is captured by the variables *yr2004* to *yr2007*. Their coefficients identify the percentage price increase relative to the year 2003. The estimates suggest that housing prices have increased by about 20% from 2003 to 2007. It is also apparent that the price increase has perceptively decreased from 2006 to 2007 compared to the price increases for the earlier years. One can presume that this slowdown in the rate of housing price inflation is a sign of the subprime mortgage problems that started to emerge with force in the latter part of 2007.

4.2 Adjusted Conditional Quantile Estimates

Based on the discussion in Zhou and Zietz (2008), the conditional quantile coefficients are likely to overstate the true dispersion of the coefficients across quantiles. Application of the adjustment factor $(1 - w)$, which is derived in section 2.3, will moderate the degree of over-dispersion. Table 6 illustrates for the quantiles from 0.1 to 0.9 how different values of the weight w affect the ability of the conditional quantile estimates to predict the unconditional quantile points. Results are provided for values of w from 0 to 1, with more detailed calculations around $w = 0.2$. A weight of $w = 0$ generates the predictions that are identified by the broken line in Figure 1, while a weight of $w = 1$ generates the thin red line. According to Table 6, either weight induces a large value for the sum of squared deviations between the unconditional quantile points (*uq*) and the predicted quantile values. Table 6 reveals that a weight of $w = 0.23$ minimizes the sum of the squared deviations over all nine quantile points. For a weight of 0.23, the conditional quantile estimates of all regressors other than the constant and the year and season dummy variables need to be multiplied by the factor $(1 - w) = (1 - 0.23) = 0.77$.

These adjusted coefficient estimates are presented in Table 7.[8] As expected, the coefficient dispersion of the quantile estimates is reduced compared to those reported in Table 4.

Since the data set covers more than one year, the question arises to what extent the flooring coefficients change over time. Since the implicit prices of a hedonic price equation are the equilibrium outcomes of demand and supply forces, changes in the implicit prices over time can be motivated either from the demand or the supply side. Demand side changes would be shifts in preferences. For example, carpet may be going out of fashion while hardwood floors are becoming more desirable. Such a change in preferences is likely reflected in a rising implicit price of hardwood flooring, except in the unrealistic case where the supply of hardwood flooring reacts instantaneously with infinite elasticity. A supply side change affecting the implicit prices of different flooring types may be induced by a commodity boom in those raw materials that are relevant for the manufacturing of flooring. As a consequence, one would expect the implicit prices of at least some flooring types to rise over time. By contrast, if the commodity boom is more general in nature and affects all building materials, then its effect is likely captured by the binary variables *yr2004* to *yr2007*, which are intended to capture general price increases in housing regardless of cause.

To test the change in implicit prices of flooring over time, interaction terms are created between all flooring variables and each of the binary variables *yr2004* to *yr2007*, which identify the years when the sales transaction took place for each home. To reduce

[8] The last six coefficients, from the variable *winter* to the regression constant, are not adjusted because a change in these variables for any quantile would imply a change for all others and, hence, for the mean.

the number of interaction terms, the different types of flooring are reduced by two. The flooring types of slate and laminate are added to the category identified as "other". This leaves seven flooring types: carpet, vinyl, wood, tile, parquet, marble, and the newly enlarged category "other". Together with four binary "year" variables, these seven categories of flooring generate 28 interaction terms. Table 8 presents the results of testing their joint statistical significance for OLS and for each quantile from 0.1 to 0.9.

Although one can reject for four quantiles the null hypothesis that there is no change in any of the flooring variables over any of the years, there is no evidence of a systematic change over time. For example, the significant p-values for quantiles 0.1 and 0.4 derive from wood flooring having a larger negative interaction term in 2007 and, for quantile 0.1, marble flooring having a larger coefficient value in 2004 and in 2006. The changes for quantile 0.8 relate to larger values for the flooring category *other* for the years 2005, 2006, and 2007. The 0.9 quantile has some significant interaction terms for individual years for carpet, parquet, marble and other flooring. Overall, the coefficients for the interaction terms are not suggestive of a trend that may be tied to changes in consumer preferences or raw material costs. As a consequence, the subsequent discussion will abstract from any changes in the implicit prices of flooring over time.

Tables 9 to 12 provide an alternative way of looking at the impact of flooring on the sales prices of single-family homes. The focus of these tables is the implicit price of alternative flooring combinations of flooring in a house. Table 9 provides a list of all the different flooring combinations that are observed and their frequency in the sample. It shows, for example, that the combination of carpet and vinyl is by far the most frequently encountered combination of flooring type. More than a third of all sales transactions from

2003 to 2007 involve houses with this type of flooring combination. The combination carpet, vinyl, and wood is observed in close to 17% of all home sales and the combination carpet, wood, and tile in approximately 14%. Together the three most frequently observed flooring combinations are quite dominant. One of them is present in roughly two thirds of all homes sold between 2003 and 2007.

Table 10 reports the implicit prices for those flooring combinations that are observed in at least 0.2% of all home sales.[9] The most popular flooring combination, carpet and vinyl, is taken to be the base category. Hence, the reported coefficients can be interpreted as percentage changes relative to the base category of flooring.[10] The last column in Table 10 lists the frequency of each flooring combination in the sample. Table 11 re-orders the variables of Table 10 in descending order of the size of the coefficients for quantile 0.1. Similarly, Table 12 presents a re-ordering of Table 10 based on the size of the coefficients for quantile 0.9.

Table 10 reveals that a flooring combination other than the dominant carpet-vinyl one can add significantly to the value of a home. For example, adding wood to a carpet-vinyl home will raise the selling price by 2% to 4.3%. Switching out the vinyl in a carpet-vinyl-wood home with tile would almost uniformly across quantiles add another 5% to the value of the home. The combination of wood and tile also captures a high price premium over the standard carpet-vinyl combination. It is interesting to note that the premium for wood-tile flooring is twice as high for a home in the lowest quantile than for a house in the highest one. This result is indicative of a general tendency. A comparison

[9] 0.2% translates into 13 home sales. Flooring combinations with fewer home sales are excluded from the regressions to prevent that unusual housing characteristics are picked up by means of the flooring variables.
[10] The results of Table 10 are retrieved from regression like those for Table 7, except for a different set of flooring variables.

of the most valued flooring combinations at quantile 0.1 (Table 11) with those at quantile 0.9 (Table 12) confirms that the type of flooring is on average far less important for determining the prices of expensive homes than the prices of inexpensive ones.

5 Conclusion

This study provides evidence on the implicit valuation of different flooring types for single family homes over the period from 2003 to 2007 in Tennessee. The empirical analysis makes use of the hedonic price model. In addition to traditional least squares estimates, the study employs conditional quantile regression, both in its traditional form, as recently applied to hedonic price function in real estate by Zietz et al. (2008) and in its modified form as suggested by Zhou and Zietz (2008). The estimates make use of a large data set of more than 6,000 sales transactions. Numerous housing variables and distance variables are used as covariates. In addition, the estimates use a spatial lag variable to account for spatial autocorrelation. The spatial lag variable is constructed along the lines of Pace et al. (2000). By including only sales of homes in the spatial weight matrix that predate the sale for each home, the spatial lag model avoids the endogeneity problem typically encountered with spatial lag variables. This greatly simplifies the estimation of the quantile regression models.

The general empirical findings are similar to those reported in Zietz et al. (2008). For many housing characteristics there are significant differences in implicit prices across the quantiles. The study shows that the value attached to individual flooring types are rather similar across home price quantiles for such common flooring types as carpet and vinyl. Wood floors, parquet and tile flooring are typically more valued at the lower end of the sales price distribution than at its upper end. Finished wood is valued the most, followed by marble and tile. Vinyl flooring lowers the price of a house on average by two percent.

There is no evidence that the implicit prices attached to different flooring types have changed over time. Preference changes are either too subtle to be captured over a five year observation period and changes in the cost of different materials appear to be very similar to those of building materials in general.

The study also provides evidence on the implicit prices attached to various combinations of flooring types. The results show that the implicit prices of numerous combinations of flooring can differ perceptively across quantiles. For example, for lower priced homes, the combination of wood and tile adds a premium of 11% compared to home with carpet-vinyl flooring. For homes at the upper end of the price distribution, the premium is just half that amount. For high-priced homes, the combination of carpet-wood-tile has the highest implicit price premium (7%). Overall, the results suggest that the particular flooring combination has less of a percentage impact on the prices of expensive homes than on the prices of inexpensive ones.

References

Basu, Sabyasachi and Thomas Thibodeau. (1998). Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics*, 17, 61-85.

Bowen, William M., B.A. Mikelbank, and D.M. Prestegaard. (2001). Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth and Change,* 32, 466-490.

Can, Ayse. (1990). The Measurement of Neighborhood Dynamics in Urban Houses Prices. *Economic Geography*, 66, 254-272.

Can, Ayse. (1992). Specification and Estimation of Hedonic Housing Price Models. *Regional Science and Urban Economics*, 22, 453-374.

Clapp, John and D. Tirtiroglu. (1994). Positive Feedback Trading and Difusion of Asset Price Changes: Evidence from Housing Transaction. *Journal of Economic Behavior and Organization,* 24, 337-355.

Dubin, R.A. (1988). Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *Review of Economics and Statistics*, 70, 466-474.

Dubin, R.A. (1992). Spatial Autocorrelation and Neighborhood Quality. *Regional Science and Urban Economics*, 22, 433-452.

Freeman, M.A, III. (1979). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. *Scandavavian Journal of Economics,* 154-173.

Halvorsen, R. and R. Palmquist. (1980). The Interpretation of Dummy Variables in Semilogrithmic Regressions. *American Economic Review*, 70, 474-475.

Holly, S., M.H. Pesaran, and T. Yamagata. (2007). A Spatio-Temporal Model of House Prices in the U.S. *University of Cambridge*, working paper.

Koenker, R. and K.F. Hallock. (2001). Quantile Regression. *Journal of Economic Perspective*, 15, 143-56.

Koenker, R. and G. Bassett. (1978). Regression Quantile. *Econometrica*, 46, 33-50.

Malpezzi, S. (2003). Hedonic Pricing Model: A Selective and Applied Review, in *Housing Economics and Public Policy: Essay in Honor of Duncan Maclennan,* T.O. Sullivan and K. Gibbs (Eds.), Blackwell.

Pace, R.K., R. Barry, O.W. Gilley, and C.F. Sirmans. (2000). A Method for Spatial-Temporal Forecasting with an Application to Real Estate Prices, *International Journal of Forecasting*, 16, 229-246.

Sirmans, G. S., D. A. Macpherson and E. N. Zietz. (2005). The Composition of Hedonic Pricing Models, Journal of Real Estate Literature, 13:1, 3-46.

Zhou, Y.X., and J. Zietz. (2008). A Modified Conditional Quantile Estimator: An Application to Hedonic Price Functions in Real Estate Economics. Department of Economics and Finance, Middle Tennessee State University.

Zietz, J., E.N. Zietz, and G.S. Sirmans. (2008). Determinants of House Prices: A Quantile Regression Approach, *Journal of Real Estate Finance and Economics*, 37, 317-333.

Table 1: Variable Definitions

Variable	Definition
lnsp	Sale price in natural logarithm (dependent variable)
Flooring types	
carpet	1 if carpet flooring is present in house, 0 otherwise
vinyl	1 if vinyl flooring is present in house, 0 otherwise
wood	1 if final wood flooring is present in house, 0 otherwise
tile	1 if tile flooring is present in house, 0 otherwise
parquet	1 if parquet flooring is present in house, 0 otherwise
marble	1 if marble flooring is present in house, 0 otherwise
slate	1 if slate flooring is present in house, 0 otherwise
lamin	1 if laminate flooring is present in house, 0 otherwise
other	1 if other flooring type is present in house, 0 otherwise
House characteristics	
sqft	Size of house in square feet, divided by 1,000
lotacre	Lot size in acres
bedroom	Number of bedrooms
bathf	Number of full bathrooms
bathh	Number of half baths
garage	Garage capacity
deck	1 if deck is present, 0 otherwise
patio	1 if patio is present, 0 otherwise
exvinyl	1 if exterior trim is made of vinyl, 0 otherwise
age	Age of the house as of the year sold
age2	Square of age, divided by 100
Neighborhood, locational characteristics, etc	
slag_lnsp	spatial lag: distance-weighted average sales price of homes in neighborhood
discenter	Distance to city center
disnash	Distance to downtown Nashville, TN
winter	1 if the house is listed in winter period, 0 otherwise
yr2004	1 if house sold in 2004, 0 otherwise
yr2005	1 if house sold in 2005, 0 otherwise
yr2006	1 if house sold in 2006, 0 otherwise
yr2007	1 if house sold in 2007, 0 otherwise

Note: Data on floor, house characteristics, and year sold come from the Middle Tennessee Regional Multiple Listing Service (MTRMLS Realtrac Inc.).

Table 2: Descriptive Statistics of Variables

	Mean	Std. Dev.	Min	Max
carpet	0.957	0.202	0	1
vinyl	0.712	0.453	0	1
wood	0.427	0.495	0	1
tile	0.295	0.456	0	1
parquet	0.068	0.251	0	1
marble	0.007	0.081	0	1
slate	0.004	0.059	0	1
lamin	0.001	0.028	0	1
other	0.035	0.184	0	1
sqft	1.842	0.689	0.6	10.358
lotacre	0.463	0.691	0	21.5
bedroom	3.179	0.518	0	6
bathf	2.083	0.488	1	5
bathh	0.304	0.478	0	3
deck	0.447	0.497	0	1
patio	0.371	0.483	0	1
garage	1.658	0.747	0	9
flcar	0.957	0.202	0	1
exvinyl	0.744	0.436	0	1
age	11.531	10.953	0	106
age2	2.529	5.793	0	112
slag_lnsp	8.232	5.452	0	13.162
discenter	3.766	2.030	0.164	25.854
disnash	27.526	5.190	8.963	44.108
winter	0.173	0.378	0	1
yr2004	0.186	0.389	0	1
yr2005	0.210	0.407	0	1
yr2006	0.227	0.419	0	1
yr2007	0.210	0.407	0	1

Note: 6,251 observations

Table 3: Percentage of Flooring in Use. by Quantiles of Sales Price

	Full	quantile									
Floor	**sample**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**
carpet	95.74	91.56	93.59	96.64	96.38	96.41	95.76	96.09	98.05	96.64	96.47
vinyl	71.24	85.94	87.34	84.00	83.46	75.98	76.14	75.08	68.18	51.20	24.40
wood	42.68	20.47	16.99	18.40	22.83	28.92	35.32	46.91	63.47	79.52	95.18
tile	29.52	7.97	11.22	15.68	17.48	22.55	24.02	27.52	37.01	51.52	81.06
parquet	6.78	4.53	7.85	11.36	9.13	8.99	8.01	7.49	6.66	2.88	0.96
marble	0.66	0.00	0.64	0.00	0.31	0.33	0.78	0.16	0.65	0.48	3.21
slate	0.35	0.31	0.48	0.00	0.16	0.65	0.16	0.49	0.49	0.48	0.32
lamin	0.08	0.00	0.00	0.00	0.16	0.33	0.16	0.00	0.16	0.00	0.00
other	3.52	2.03	2.72	4.00	4.88	4.41	4.71	5.54	3.90	1.60	1.44
observ.	6,251	640	624	625	635	612	637	614	616	625	623

Table 4: Coefficient Estimates, by OLS and Conditional Quantile Regression, 6,251 Observations

Variable	OLS	quantile 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Flooring types										
carpet	0.0123	0.0154	0.0222	0.0132	0.0128	0.0048	0.0092	0.0071	0.0088	-0.0001
vinyl	-0.0217	-0.0207	-0.0197	-0.0169	-0.0215	-0.0206	-0.0208	-0.0197	-0.0207	-0.0116
wood	0.0610	0.0617	0.0504	0.0447	0.0436	0.0413	0.0429	0.0399	0.0357	0.0397
tile	0.0404	0.0434	0.0364	0.0360	0.0307	0.0303	0.0240	0.0239	0.0241	0.0338
parquet	0.0123	0.0279	0.0149	0.0139	0.0068	0.0061	0.0039	-0.0009	-0.0018	-0.0006
marble	0.0485	-0.0017	0.0360	0.0383	0.0341	0.0250	0.0179	0.0396	0.0281	0.0061
slate	-0.0123	-0.0537	0.0139	0.0159	0.0140	-0.0045	-0.0097	-0.0258	-0.0249	0.0123
lamin	0.0414	0.0213	0.0485	0.0258	0.0593	0.0389	0.0174	0.0274	0.0083	0.0559
other	0.0073	0.0053	0.0151	0.0138	0.0067	0.0046	0.0024	0.0032	0.0077	0.0055
House characteristics										
sqft	0.3464	0.3158	0.3392	0.3581	0.3687	0.3812	0.3844	0.3962	0.4028	0.4077
lotacre	0.0282	0.0215	0.0331	0.0313	0.0301	0.0304	0.0317	0.0349	0.0440	0.0507
bedroom	-0.0064	0.0064	-0.0012	-0.0089	-0.0130	-0.0175	-0.0157	-0.0158	-0.0141	-0.0239
bathf	0.0171	0.0031	0.0039	0.0033	0.0074	0.0063	0.0139	0.0135	0.0202	0.0350
bathh	0.0060	-0.0045	-0.0011	-0.0051	-0.0114	-0.0128	-0.0086	-0.0103	-0.0038	0.0010
deck	0.0102	0.0316	0.0175	0.0148	0.0115	0.0092	0.0082	0.0028	-0.0006	-0.0080
patio	0.0080	0.0194	0.0092	0.0050	0.0056	0.0069	0.0055	0.0015	-0.0002	-0.0054
garage	0.0599	0.0635	0.0576	0.0561	0.0557	0.0545	0.0549	0.0560	0.0549	0.0556
exvinyl	-0.0085	0.0035	-0.0035	-0.0074	-0.0106	-0.0088	-0.0088	-0.0095	-0.0106	-0.0151
age	-0.0033	-0.0028	-0.0033	-0.0032	-0.0033	-0.0031	-0.0031	-0.0027	-0.0031	-0.0031
age2	-0.0041	-0.0110	-0.0067	-0.0051	-0.0036	-0.0030	-0.0019	-0.0019	-0.0006	0.0005

Table 4: Coefficient Estimates, by OLS and Conditional Quantile Regression, 6,251 Observations (continued)

Variable	OLS	quantile 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Neighborhood variable, locational characteristics, etc.										
slag_lnsp	0.0004	0.0004	0.0004	0.0002	0.0005	0.0002	0.0001	0.0000	0.0000	0.0002
discenter	-0.0024	-0.0040	-0.0031	-0.0027	-0.0021	-0.0015	-0.0008	-0.0008	-0.0006	-0.0015
disnash	-0.0028	-0.0026	-0.0033	-0.0031	-0.0032	-0.0033	-0.0034	-0.0032	-0.0031	-0.0031
winter	-0.0198	-0.0283	-0.0253	-0.0248	-0.0201	-0.0187	-0.0151	-0.0132	-0.0129	-0.0041
yr2004	0.0549	0.0634	0.0556	0.0546	0.0547	0.0531	0.0534	0.0533	0.0569	0.0623
yr2005	0.1120	0.1164	0.1101	0.1077	0.1116	0.1104	0.1097	0.1103	0.1123	0.1174
yr2006	0.1695	0.1751	0.1750	0.1756	0.1745	0.1706	0.1686	0.1685	0.1689	0.1702
yr2007	0.2031	0.2131	0.2102	0.2125	0.2129	0.2073	0.2043	0.2016	0.2000	0.1984
Constant	11.1319	11.0353	11.0988	11.1305	11.1479	11.1750	11.1612	11.1576	11.1464	11.1697
R^2	0.9003	0.6197	0.6485	0.6747	0.6870	0.7177	0.7355	0.7502	0.7648	0.7798

Notes: The R^2 listed for the quantile regressions are pseudo R^2.

Table 5: P-values for Coefficients of Table 4, OLS and Conditional Quantile Regression

Variable	OLS	quantile 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Flooring types										
carpet	0.191	0.169	0.030	0.043	0.052	0.499	0.105	0.207	0.203	0.994
vinyl	0	0.001	0	0	0	0	0	0	0	0.007
wood	0	0	0	0	0	0	0	0	0	0
tile	0	0	0	0	0	0	0	0	0	0
parquet	0.008	0.002	0.066	0.007	0.196	0.283	0.392	0.833	0.746	0.928
marble	0.021	0.950	0.140	0.014	0.029	0.142	0.190	0.004	0.093	0.767
slate	0.742	0.128	0.666	0.441	0.505	0.846	0.590	0.153	0.261	0.649
lamin	0.274	0.375	0.480	0.499	0.161	0.377	0.635	0.413	0.860	0.002
other	0.382	0.649	0.166	0.049	0.345	0.547	0.697	0.597	0.300	0.544
House characteristics										
sqft	0	0	0	0	0	0	0	0	0	0
lotacre	0	0	0	0	0	0	0	0	0	0
bedroom	0.367	0.358	0.844	0.013	0	0	0	0	0	0
bathf	0.008	0.700	0.596	0.471	0.095	0.171	0	0	0	0
bathh	0.303	0.495	0.841	0.147	0.001	0.001	0.004	0.001	0.286	0.808
deck	0.003	0	0	0	0	0.008	0.003	0.314	0.858	0.049
patio	0.026	0	0.065	0.120	0.082	0.049	0.051	0.600	0.959	0.196
garage	0	0	0	0	0	0	0	0	0	0
exvinyl	0.021	0.505	0.472	0.017	0.001	0.011	0.001	0.001	0.002	0
age	0	0	0	0	0	0	0	0	0	0
age2	0.037	0	0	0	0	0	0	0	0.385	0.606

Table 5: P-values for Coefficients of Table 4, OLS and Conditional Quantile Regression (continued)

Variable	OLS	quantile 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Neighborhood, locational characteristics, and others										
slag_lnsp	0.154	0.290	0.275	0.303	0.051	0.370	0.599	0.835	0.912	0.488
discenter	0.005	0.002	0.004	0.000	0.001	0.038	0.140	0.174	0.350	0.070
disnash	0	0	0	0	0	0	0	0	0	0
winter	0	0	0	0	0	0	0	0	0	0.348
yr2004	0	0	0	0	0	0	0	0	0	0
yr2005	0	0	0	0	0	0	0	0	0	0
yr2006	0	0	0	0	0	0	0	0	0	0
yr2007	0	0	0	0	0	0	0	0	0	0
Constant	0	0	0	0	0	0	0	0	0	0

Notes: A p-value of 0.000 indicates statistical significance at much better than the one percent level.

Table 6: Selection of Adjustment Weight *w*

	Quantile									
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**ssqdev**
uq	11.5521	11.6527	11.7272	11.7906	11.8565	11.9512	12.0553	12.1704	12.3712	
weight										
0	11.4874	11.6126	11.7048	11.7806	11.8531	11.9616	12.0872	12.2340	12.4817	0.02379
0.16	11.5336	11.6478	11.7308	11.7989	11.8640	11.9592	12.0691	12.1973	12.4128	0.00321
0.17	11.5365	11.6500	11.7324	11.8001	11.8646	11.9590	12.0679	12.1950	12.4084	0.00265
0.18	11.5394	11.6522	11.7341	11.8012	11.8653	11.9589	12.0668	12.1927	12.4041	0.00218
0.19	11.5423	11.6544	11.7357	11.8024	11.8660	11.9587	12.0657	12.1904	12.3998	0.00179
0.2	11.5452	11.6566	11.7373	11.8035	11.8667	11.9585	12.0645	12.1881	12.3955	0.00148
0.21	11.5481	11.6588	11.7389	11.8047	11.8674	11.9584	12.0634	12.1858	12.3912	0.00127
0.22	11.5509	11.6610	11.7406	11.8058	11.8681	11.9582	12.0623	12.1836	12.3869	0.00113
0.23	11.5538	11.6632	11.7422	11.8070	11.8687	11.9581	12.0611	12.1813	12.3826	0.00109
0.24	11.5567	11.6654	11.7438	11.8081	11.8694	11.9579	12.0600	12.1790	12.3783	0.00112
0.25	11.5596	11.6676	11.7455	11.8093	11.8701	11.9578	12.0589	12.1767	12.3740	0.00125
0.26	11.5625	11.6698	11.7471	11.8104	11.8708	11.9576	12.0577	12.1744	12.3697	0.00146
0.27	11.5654	11.6720	11.7487	11.8115	11.8715	11.9575	12.0566	12.1721	12.3654	0.00175
1	11.7763	11.8326	11.8675	11.8953	11.9213	11.9465	11.9739	12.0048	12.0510	0.25406

Notes: *uq* identifies the unconditional quantile values; *ssqdev* is the sum of the squared deviations, which are calculated for each value of *w* by summing the squared deviations between *uq* and the predicted quantile value at each quantile point.

Table 7: Coefficient Estimates, OLS and Adjusted Conditional Quantile Regression, 6,251 Observations

		quantile								
	OLS	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Flooring types										
carpet	0.0095	0.0118	0.0171	0.0102	0.0099	0.0037	0.0071	0.0055	0.0067	-0.0001
vinyl	-0.0167	-0.0160	-0.0152	-0.0130	-0.0165	-0.0158	-0.0160	-0.0152	-0.0160	-0.0089
wood	0.0470	0.0475	0.0388	0.0344	0.0336	0.0318	0.0330	0.0307	0.0275	0.0306
tile	0.0311	0.0334	0.0280	0.0278	0.0236	0.0234	0.0185	0.0184	0.0186	0.0260
parquet	0.0094	0.0215	0.0115	0.0107	0.0052	0.0047	0.0030	-0.0007	-0.0014	-0.0005
marble	0.0373	-0.0013	0.0278	0.0295	0.0263	0.0193	0.0138	0.0305	0.0216	0.0047
slate	-0.0095	-0.0414	0.0107	0.0122	0.0108	-0.0034	-0.0074	-0.0199	-0.0192	0.0094
lamin	0.0319	0.0164	0.0373	0.0199	0.0456	0.0300	0.0134	0.0211	0.0064	0.0430
other	0.0056	0.0041	0.0117	0.0106	0.0051	0.0035	0.0018	0.0025	0.0059	0.0042
House characteristics										
sqft	0.2668	0.2432	0.2612	0.2758	0.2839	0.2935	0.2960	0.3050	0.3101	0.3139
lotacre	0.0217	0.0166	0.0255	0.0241	0.0232	0.0234	0.0244	0.0269	0.0339	0.0390
bedroom	-0.0049	0.0050	-0.0009	-0.0069	-0.0100	-0.0135	-0.0121	-0.0122	-0.0109	-0.0184
bathf	0.0131	0.0024	0.0030	0.0025	0.0057	0.0049	0.0107	0.0104	0.0156	0.0269
bathh	0.0046	-0.0034	-0.0009	-0.0039	-0.0088	-0.0098	-0.0066	-0.0079	-0.0030	0.0008
deck	0.0078	0.0243	0.0135	0.0114	0.0089	0.0071	0.0063	0.0022	-0.0005	-0.0062
patio	0.0062	0.0149	0.0071	0.0038	0.0043	0.0053	0.0042	0.0011	-0.0001	-0.0041
garage	0.0461	0.0489	0.0444	0.0432	0.0429	0.0420	0.0423	0.0432	0.0423	0.0428
exvinyl	-0.0066	0.0027	-0.0027	-0.0057	-0.0081	-0.0067	-0.0067	-0.0073	-0.0082	-0.0116
age	-0.0026	-0.0022	-0.0026	-0.0025	-0.0025	-0.0024	-0.0024	-0.0021	-0.0024	-0.0024
age2	-0.0032	-0.0085	-0.0052	-0.0039	-0.0028	-0.0023	-0.0015	-0.0015	-0.0004	0.0004
Neighborhood, locational characteristics, etc.										
slag_lnsp	0.0003	0.0003	0.0003	0.0002	0.0004	0.0002	0.0001	0.0000	0.0000	0.0002
discenter	-0.0018	-0.0031	-0.0024	-0.0021	-0.0016	-0.0011	-0.0006	-0.0006	-0.0005	-0.0011
disnash	-0.0021	-0.0020	-0.0025	-0.0024	-0.0025	-0.0026	-0.0026	-0.0024	-0.0024	-0.0024
winter	-0.0198	-0.0283	-0.0253	-0.0248	-0.0201	-0.0187	-0.0151	-0.0132	-0.0129	-0.0041
yr2004	0.0549	0.0634	0.0556	0.0546	0.0547	0.0531	0.0534	0.0533	0.0569	0.0623
yr2005	0.1120	0.1164	0.1101	0.1077	0.1116	0.1104	0.1097	0.1103	0.1123	0.1174
yr2006	0.1695	0.1751	0.1750	0.1756	0.1745	0.1706	0.1686	0.1685	0.1689	0.1702
yr2007	0.2031	0.2131	0.2102	0.2125	0.2129	0.2073	0.2043	0.2016	0.2000	0.1984
Constant	11.1319	11.0353	11.0988	11.1305	11.1479	11.1750	11.1612	11.1576	11.1464	11.1697

Notes: The adjustment factor is (1-*w*) = 0.77. No adjustment factor is applied to the last six variables because they are not house-specific variables. The p-values of Table 5 apply.

Table 8: Test Results of Adding Time Interaction Terms for all Flooring Types

		quantiles								
	OLS	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
p-value	0.1736	0.0001	0.1298	0.0748	0.0253	0.1583	0.1924	0.0899	0.0216	0.0000

Notes: The p-values test the null hypotheses that the interaction terms of all flooring variables with the variables *yr2004* to *yr2007* are jointly equal to zero. There are 28 such interaction variables in each regression. The flooring types slate and laminate are adding to the category *other* for a total of seven flooring categories.

Table 9: Observed Flooring Combinations, Ordered by Percentage of Use, 6,251 Observations

	Percent	carpet	vinyl	wood	tile	parquet	marble	other
flcv	0.34890	1	1	0	0	0	0	0
flcvw	0.16829	1	1	1	0	0	0	0
flcwt	0.13774	1	0	1	1	0	0	0
flcvp	0.05135	1	1	0	0	1	0	0
flcvwt	0.04799	1	1	1	1	0	0	0
flc	0.04399	1	0	0	0	0	0	0
flcvt	0.04127	1	1	0	1	0	0	0
flct	0.03343	1	0	0	1	0	0	0
flcw	0.03120	1	0	1	0	0	0	0
flcvo	0.01696	1	1	0	0	0	0	1
flvw	0.00864	0	1	1	0	0	0	0
flv	0.00816	0	1	0	0	0	0	0
flw	0.00768	0	0	1	0	0	0	0
flwt	0.00640	0	0	1	1	0	0	0
flvwt	0.00496	0	1	1	1	0	0	0
flco	0.00480	1	0	0	0	0	0	1
flcvtp	0.00400	1	1	0	1	1	0	0
flcto	0.00352	1	0	0	1	0	0	1
flcvto	0.00336	1	1	0	1	0	0	1
flcvwp	0.00320	1	1	1	0	1	0	0
flctp	0.00288	1	0	0	1	1	0	0
flcp	0.00256	1	0	0	0	1	0	0
flcwtm	0.00208	1	0	1	1	0	1	0
flt	0.00144	0	0	0	1	0	0	0
flcwto	0.00144	1	0	1	1	0	0	1
flcwtp	0.00128	1	0	1	1	1	0	0
flcwo	0.00128	1	0	1	0	0	0	1
flo	0.00112	0	0	0	0	0	0	1
flcvm	0.00112	1	1	0	0	0	1	0
flvt	0.00096	0	1	0	1	0	0	0
flcvwm	0.00096	1	1	1	0	0	1	0
flctm	0.00096	1	0	0	1	0	1	0
flcwm	0.00080	1	0	1	0	0	1	0
flcwp	0.00064	1	0	1	0	1	0	0
flcvwo	0.00064	1	1	1	0	0	0	1
flwo	0.00048	0	0	1	0	0	0	1
flp	0.00032	0	0	0	0	1	0	0
flvp	0.00032	0	1	0	0	1	0	0
flwp	0.00032	0	0	1	0	1	0	0
flwtm	0.00032	0	0	1	1	0	1	0
flvo	0.00032	0	1	0	0	0	0	1
flto	0.00032	0	0	0	1	0	0	1
flvto	0.00032	0	1	0	1	0	0	1
flcvpo	0.00032	1	1	0	0	1	0	1

Notes: Six flooring combinations, which are used only in a single house, are not listed.

Table 10: Flooring Coefficients, OLS and Adjusted Quantile Regression, 6,251 Observations

		quantile									
	OLS	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**Percent**
flcvw	0.039	0.043	0.031	0.030	0.028	0.028	0.027	0.026	0.020	0.026	0.168
flcwt	0.098	0.103	0.086	0.084	0.080	0.073	0.069	0.071	0.072	0.070	0.138
flcvp	0.007	0.020	0.013	0.011	0.007	0.004	0.001	-0.001	-0.002	-0.001	0.051
flcvwt	0.067	0.064	0.055	0.049	0.047	0.048	0.046	0.051	0.051	0.054	0.048
flc	0.000	-0.010	-0.013	-0.002	0.003	0.004	0.007	0.008	0.006	0.003	0.044
flcvt	0.017	0.025	0.015	0.026	0.017	0.018	0.012	0.012	0.010	0.016	0.041
flct	0.039	0.057	0.043	0.037	0.034	0.035	0.030	0.026	0.020	0.013	0.033
flcw	0.047	0.056	0.046	0.045	0.041	0.041	0.040	0.041	0.035	0.031	0.031
flcvo	0.014	0.018	0.021	0.016	0.013	0.007	0.006	0.012	0.008	0.014	0.017
flvw	0.043	0.023	0.029	0.023	0.020	0.042	0.021	0.022	0.018	0.023	0.009
flv	-0.012	-0.011	-0.018	-0.024	-0.019	-0.023	-0.011	-0.011	-0.014	0.015	0.008
flw	0.062	0.070	0.065	0.058	0.066	0.056	0.055	0.043	0.040	0.035	0.008
flwt	0.092	0.109	0.078	0.087	0.080	0.064	0.053	0.064	0.058	0.053	0.006
flvwt	0.046	0.072	0.036	0.043	0.033	0.024	0.015	0.017	0.002	0.015	0.005
flco	0.029	0.018	0.032	0.016	0.009	0.010	0.016	0.013	0.007	0.008	0.005
flcvtp	0.038	0.064	0.051	0.031	0.016	0.011	0.017	0.004	0.006	0.031	0.004
flcto	0.024	0.045	0.016	0.026	0.028	0.013	0.016	0.006	0.013	0.004	0.004
flcvto	0.038	0.061	0.036	0.035	0.032	0.026	0.027	0.018	0.018	0.009	0.003
flcvwp	0.034	0.056	0.029	0.034	0.031	0.014	0.005	0.018	0.010	-0.002	0.003
flctp	0.024	0.018	0.031	0.022	0.017	0.015	0.005	0.018	0.022	0.052	0.003
flcp	0.045	0.054	0.006	0.023	0.030	0.022	0.042	0.047	0.036	0.059	0.003
flcwtm	0.116	0.070	0.037	0.034	0.145	0.118	0.107	0.104	0.081	0.062	0.002

Notes: The adjustment factor is $(1-w) = 0.77$. Shaded fields identify statistical significance at the 5% level or better.

Table 11: Flooring Coefficients, Variables Ordered by Size of Coefficients of Quantile 0.1

	OLS	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Percent
flwt	0.092	0.109	0.078	0.087	0.080	0.064	0.053	0.064	0.058	0.053	0.006
flcwt	0.098	0.103	0.086	0.084	0.080	0.073	0.069	0.071	0.072	0.070	0.138
flvwt	0.046	0.072	0.036	0.043	0.033	0.024	0.015	0.017	0.002	0.015	0.005
flcwtm	0.116	0.070	0.037	0.034	0.145	0.118	0.107	0.104	0.081	0.062	0.002
flw	0.062	0.070	0.065	0.058	0.066	0.056	0.055	0.043	0.040	0.035	0.008
flcvtp	0.038	0.064	0.051	0.031	0.016	0.011	0.017	0.004	0.006	0.031	0.004
flcvwt	0.067	0.064	0.055	0.049	0.047	0.048	0.046	0.051	0.051	0.054	0.048
flcvto	0.038	0.061	0.036	0.035	0.032	0.026	0.027	0.018	0.018	0.009	0.003
flct	0.039	0.057	0.043	0.037	0.034	0.035	0.030	0.026	0.020	0.013	0.033
flcw	0.047	0.056	0.046	0.045	0.041	0.041	0.040	0.041	0.035	0.031	0.031
flcvwp	0.034	0.056	0.029	0.034	0.031	0.014	0.005	0.018	0.010	-0.002	0.003
flcp	0.045	0.054	0.006	0.023	0.030	0.022	0.042	0.047	0.036	0.059	0.003
flcto	0.024	0.045	0.016	0.026	0.028	0.013	0.016	0.006	0.013	0.004	0.004
flcvw	0.039	0.043	0.031	0.030	0.028	0.028	0.027	0.026	0.020	0.026	0.168
flcvt	0.017	0.025	0.015	0.026	0.017	0.018	0.012	0.012	0.010	0.016	0.041
flvw	0.043	0.023	0.029	0.023	0.020	0.042	0.021	0.022	0.018	0.023	0.009
flcvp	0.007	0.020	0.013	0.011	0.007	0.004	0.001	-0.001	-0.002	-0.001	0.051
flco	0.029	0.018	0.032	0.016	0.009	0.010	0.016	0.013	0.007	0.008	0.005
flctp	0.024	0.018	0.031	0.022	0.017	0.015	0.005	0.018	0.022	0.052	0.003
flcvo	0.014	0.018	0.021	0.016	0.013	0.007	0.006	0.012	0.008	0.014	0.017
flc	0.000	-0.010	-0.013	-0.002	0.003	0.004	0.007	0.008	0.006	0.003	0.044
flv	-0.012	-0.011	-0.018	-0.024	-0.019	-0.023	-0.011	-0.011	-0.014	0.015	0.008

Notes: The adjustment factor is (1-*w*) = 0.77. Shaded fields identify statistical significance at the 5% level or better.

Table 12: Flooring Coefficients, Variables Ordered by Size of Coefficients of Quantile 0.9

	OLS	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Percent
flcwt	0.098	0.103	0.086	0.084	0.080	0.073	0.069	0.071	0.072	0.070	0.138
flcwtm	0.116	0.070	0.037	0.034	0.145	0.118	0.107	0.104	0.081	0.062	0.002
flcp	0.045	0.054	0.006	0.023	0.030	0.022	0.042	0.047	0.036	0.059	0.003
flcvwt	0.067	0.064	0.055	0.049	0.047	0.048	0.046	0.051	0.051	0.054	0.048
flwt	0.092	0.109	0.078	0.087	0.080	0.064	0.053	0.064	0.058	0.053	0.006
flctp	0.024	0.018	0.031	0.022	0.017	0.015	0.005	0.018	0.022	0.052	0.003
flw	0.062	0.070	0.065	0.058	0.066	0.056	0.055	0.043	0.040	0.035	0.008
flcw	0.047	0.056	0.046	0.045	0.041	0.041	0.040	0.041	0.035	0.031	0.031
flcvtp	0.038	0.064	0.051	0.031	0.016	0.011	0.017	0.004	0.006	0.031	0.004
flcvw	0.039	0.043	0.031	0.030	0.028	0.028	0.027	0.026	0.020	0.026	0.168
flvw	0.043	0.023	0.029	0.023	0.020	0.042	0.021	0.022	0.018	0.023	0.009
flcvt	0.017	0.025	0.015	0.026	0.017	0.018	0.012	0.012	0.010	0.016	0.041
flvwt	0.046	0.072	0.036	0.043	0.033	0.024	0.015	0.017	0.002	0.015	0.005
flv	-0.012	-0.011	-0.018	-0.024	-0.019	-0.023	-0.011	-0.011	-0.014	0.015	0.008
flcvo	0.014	0.018	0.021	0.016	0.013	0.007	0.006	0.012	0.008	0.014	0.017
flct	0.039	0.057	0.043	0.037	0.034	0.035	0.030	0.026	0.020	0.013	0.033
flcvto	0.038	0.061	0.036	0.035	0.032	0.026	0.027	0.018	0.018	0.009	0.003
flco	0.029	0.018	0.032	0.016	0.009	0.010	0.016	0.013	0.007	0.008	0.005
flcto	0.024	0.045	0.016	0.026	0.028	0.013	0.016	0.006	0.013	0.004	0.004
flc	0.000	-0.010	-0.013	-0.002	0.003	0.004	0.007	0.008	0.006	0.003	0.044
flcvp	0.007	0.020	0.013	0.011	0.007	0.004	0.001	-0.001	-0.002	-0.001	0.051
flcvwp	0.034	0.056	0.029	0.034	0.031	0.014	0.005	0.018	0.010	-0.002	0.003

Notes: The adjustment factor is (1-*w*) = 0.77. Shaded fields identify statistical significance at the 5% level or better.

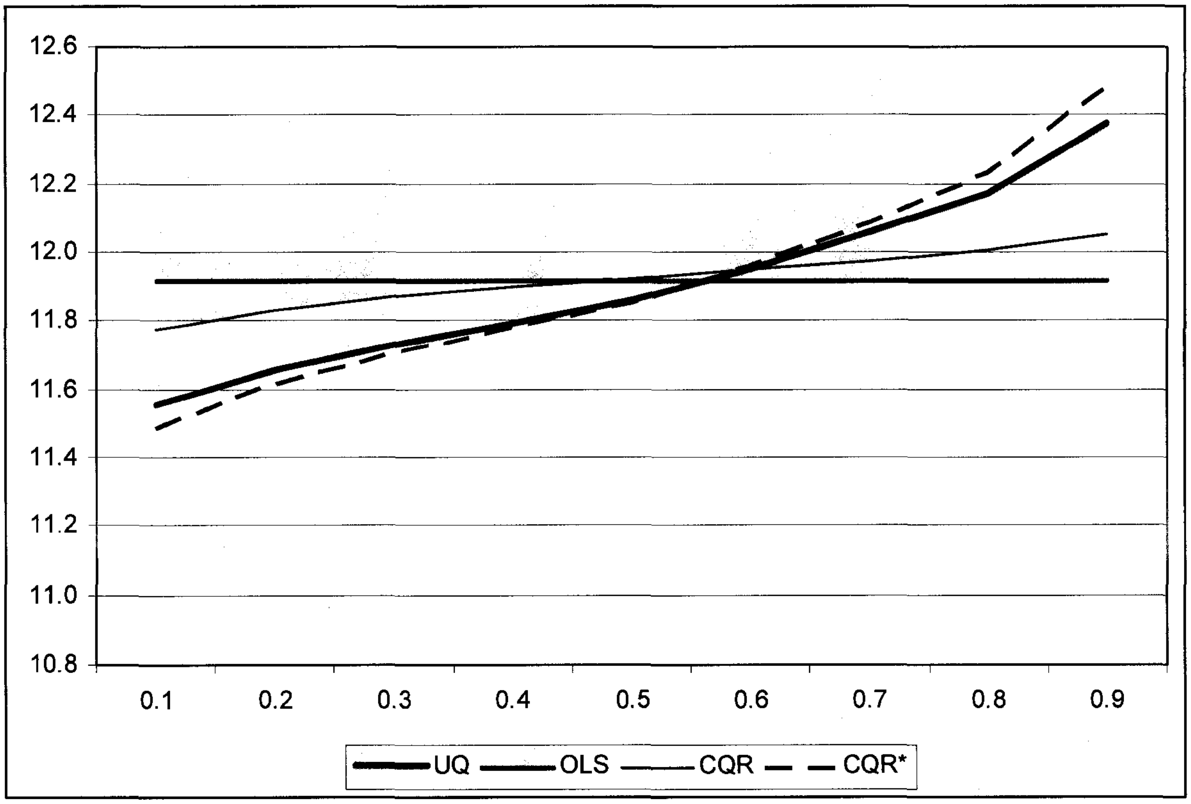


Figure 1: Unconditional Quantiles, Unconditional Mean, Predicted Quantiles Evaluated At The Mean of The Regressors, and Predicted Quantiles Evaluated At Smoothed Values of The Regressors

AN EMPIRICAL STUDY OF TIME-ON-MARKET USING QUANTILE REGRESSION

Abstract

This paper examines the determinants of Time-On-Market (TOM) for single-family homes in a Tennessee county from the years 2003 to 2007. In contrast to the previous literature on TOM, which typically uses duration/survival models, quantile regression is applied in this paper. This new approach allows for variation in the coefficients across the quantiles of the sales duration. This can increase the plausibility of estimation results and allow for new insights into the determinants of TOM for homes that are new on the market versus those that have been on the market for a long time. The results are contrasted with those obtained from the conventional Cox proportional hazards model. As in previous studies, there is evidence that several of the physical attributes of homes, such as square footage, lot size, and indicators of unusualness have a significant impact on TOM. Similarly important is the initial listing price relative to earlier sales prices in the neighborhood and school zoning. Seasonal effects and variations over time that result from changes in the macroeconomic environment are also found to have an impact. Perceptible differences are found to exist across quantiles for the coefficients of several of the variables. This is consistent with the finding that the proportionality assumptions of the traditional hazards models are rejected for all models. Although the estimated models allow for spatial correlation through a spatial lag model no spatial clustering can be identified.

1 Introduction

Time-On-Market (TOM), or the sales duration, is defined as the length of time between the date when the house is listed for sale on the market and the date when the house is sold. For the past three decades TOM has captured the interest of both practitioners and academic researchers. For example, numerous studies explore the interaction of price and TOM (Cubin, 1974; Miller, 1978; Kang and Gardner, 1989; Horowitz, 1992; Asabere and Huffman, 1993; Asabere, Huffman, and Mehdian, 1993; Anglin, Rutherford, and Springer, 2003; Knight, 2002). Other studies focus on agency and commission related issues (Miceli, 1991; Yavas, 1993; Yang and Yavas, 1995; Jud et al., 1996; Miceli et al., 2000). Several studies examine the relationship between sales duration and the behavior of buyers (Zumpano et al., 1996; D'urso, 2002), atypical house characteristics (Haurin, 1988; Robinson and Waller, 2005), seller motivation (Glower, Haurin, and Hendershott, 1998), and macroeconomic factors (Kalra and Chan, 1994; Yavas and Yang, 1995).

A distinguishing feature of housing markets is that houses are heterogeneous. That is the major reason why some earlier literature divides the housing market into homogeneous submarkets and analyzes TOM within these submarkets. For example, Yavas and Yang (1995) examine TOM along the distribution of list prices. A similar approach is employed by Belkin et al. (1976) and Kang and Gardner (1989). Leung et al. (2002) split their 7-year data into 14 half-year sub-groups. Segmentation methods, however, can be problematic because of the loss of observations due to sample truncation.

Since the late 1990s, quantile regression has seen several applications in duration analysis (Koenker and Basset, 1978; Horowitz and Neumann, 1987; Koenker and Bilias, 2001; Koenker and Hallock, 2001; Koenker and Geling, 2001; Koenker and Xiao, 2002; Portnoy, 2003; Fitzenberger and Wilke, 2005), although not yet in real estate economics. The purpose of this study is to examine to what extent quantile regression can be a useful alternative to established methods for the analysis of time on the market.

There are numerous reasons to seriously consider quantile regression instead of the proportional hazards model. First, quantile regression allows the impact of variables that determine TOM to vary by quantile. This offers similar advantages than traditional segmentation methods but without discarding sample observations. It can identify effects that arise only at certain quantiles of the distribution, such as at the extreme ends, or it can pick up shifts in the location or changes in the shape of the distribution. Such findings are beyond the reach of conventional methods. Second, quantile regression is much less restrictive in its assumptions and far closer to traditional least squares regression than duration models. This is likely to be an attractive feature for practical applications because it significantly simplifies estimation and interpretation and offers a viable estimation alternative if the restrictive assumptions of duration models are violated. Third, the basic motivation for using duration models, its ability to deal with censored data, does not apply to the typical TOM investigation in real estate, which is based on transactions data.[1]

The empirical analysis is based on more than 5,000 sales transactions of single-family homes in a Tennessee county from 2003 to 2007. Listing and closing dates are available for all observations. The paper incorporates a variety of factors including numerous physical attributes of each house, location including high school zone,

[1] Censored Quantile Regression can be used to take account of right censored records, which are typical in duration data outside of real estate; see Powell (1984, 1986) and Portnoy (2003).

measures of unusualness, and neighborhood effects to identify the determinants of the speed with which a house sells.[2]

The plan of this paper is as follows. The next section introduces some methodological background information on both the traditional duration model and the quantile regression approach to duration analysis. Section 3 describes the data set. Section 4 discusses the estimation results. The last section concludes.

2 Methodology

2.1 Cox Proportional Hazards Model

The Cox proportional hazards model is one of the most commonly used duration models, first introduced by Cox (1972). The general form of the duration model is given as

$$h(t \mid \mathbf{x}) = h_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}), \tag{1}$$

where the dependent variable is the hazard function of TOM and where the right-hand side is split up into a product of two independent terms: (a) the unknown baseline hazard function $h_0(t)$, which depends only on t and is identical for all observations, and (b) the scale factor $\exp(\mathbf{x}'\boldsymbol{\beta})$, which depends on the covariates $\mathbf{x}$ with associated coefficients $\boldsymbol{\beta}$,

[2] School zone has been identified as a potential determinant of locational choice by Hayes and Taylor (1996) and Walden (1990).

but does not depend on t.[3] This independence property makes it possible to estimate the coefficients $\boldsymbol{\beta}$ without specifying a functional form for $h_0(t)$.

The hazard function can be thought of as the instantaneous rate of a sale at time t, given that the house has not sold until t.[4] It can be formalized as

$$h(t) = \frac{dF(t)/dt}{1-F(t)},$$

where the numerator is the derivative with respect to t of the probability $F(t)$ that the house will sell before or at time t and where $1-F(t)$ is the probability that the house will *not* sell before time t,

In analogy to the linear regression model, the coefficients of the hazard function can be given the interpretation of partial derivatives (Kiefer 1988). A positive coefficient means that TOM will be shortened, a negative sign that TOM will be longer. For each coefficient (β_i), a hazard ratio can be calculated as the exponent of the coefficient $\left(e^{\beta_i}\right)$. Such a hazard ratio measures the relative rate of sales. For example, a coefficient value of -0.04 for a binary variable that identifies the initial listing of a house during the winter months translates into a hazard ratio of 0.96 $(\exp(-0.04)=0.96)$. It implies that the estimated relative rate of sales of all houses listed in the winter is 96% of all houses that were not listed in the winter. In other words, a winter listing lowers the expected rate of sale by 4%. This ratio is assumed to be constant over time. The hazard ratio for a continuous variable can be interpreted in a manner similar to the one for a binary variable. For example, a hazard ratio of 0.90 on a variable that measures square footage in units of

[3] Unlike the Cox model, the Weibull model specifies a specific parametric functional form for the baseline hazard: $h_0(t) = \alpha t^{\alpha-1}$, which adds one parameter (α) that needs to be estimated.

[4] Note that the hazard rate is not a probability but a probability rate. Its value can exceed unity.

1,000 square feet implies that a house with 1,000 additional square feet would be expected to sell at a 10% lower rate than houses without the additional 1,000 square feet.

2.2 *Quantile Regression for Duration Data*

Analogous to least squares, quantile regression models the time between the listing of a house and its sale as a linear function of a set of covariates.[5] However, in contrast to least squares and the proportional hazards model discussed in the last section, a separate regression is estimated for each desired quantile point of the distribution of TOM. The coefficient of a covariate at the q^{th} quantile point of TOM can be interpreted as approximately equal to the change in the q^{th} quantile value of TOM in response to a unit change in the value of the corresponding covariate at that quantile.

We model the conditional quantile functions of the logarithm of TOM as linear in the observed covariates. Given $y = \ln(TOM)$, the model is equivalent to:

$$Q_q(y \mid \mathbf{x}) = \sum_{j=0}^{k} b_{j,q} x_{j,i} \text{ for } q \in (0,1), \tag{2}$$

where $\mathbf{x}$ is a $(k+1)\times 1$ vector of covariates, $x_{j,i}$ is covariate j at observation i, and $b_{j,q}$ is the parameter for covariate j at the q^{th} quantile. The coefficient $b_{j,q}$ is defined as the partial derivative, with respect to covariate j, of the predicted response variable at q^{th} quantile.

[5] The relationship can be nonlinear by incorporating the quadratic forms of covariates.

For continuous variable x_j, $\beta_{q,j}$ measures the percentage change in the q^{th} quantile of TOM associated with a unit increase in x_j; if x_j is a binary variable, $\beta_{q,j}$ provides the percentage change in the q^{th} quantile of TOM resulting from a shift in x_j from 0 to 1, holding all other variables fixed at their given values. For both types of variables, a negative coefficient signals a reduction in TOM and, hence, a faster sale and a positive coefficient indicates an increase in TOM or a slower sale.

Using a logarithmic transformation of the response variable is suggested by Koenker and Bilias (2001). It is motivated by the desire to linearize the parametric specification and to ease interpretation.

Analogous to least squares regression, the estimation of quantile regression involves a minimization problem. For quantile regressions the objective is to minimize the asymmetrically weighted absolute deviation from the q^{th} predicted or conditional quantile value ($y_q = \sum_{j=0}^{k} b_{j,q} x_{j,i}$) with respect to the coefficient set,

$$\min_{\{b_{j,q}\}_{j=0}^{k}} \sum_i h_i \left| y_i - \sum_{j=0}^{k} b_{j,q} x_{j,i} \right| \text{ for } q \in (0,1),$$

where h represents the weight and where

$$h_i = \begin{cases} 2q & \text{for } y_i - \sum_{j=0}^{k} b_{j,q} x_{j,i} > 0 \\ 2-2q & \text{for } y_i - \sum_{j=0}^{k} b_{j,q} x_{j,i} \leq 0 \end{cases}.$$

At the median $(q = 0.5)$, the weight is symmetric; at other quantiles, the weight is asymmetric. A heavier weight is imposed as quantiles deviate from the median.

3 Data

The estimates are based on sales transactions data of single-family homes in Rutherford County, TN, over the years 2003 to 2007.[6] For each sales transaction, the sample includes the time on the market in days, the original listing price, physical attributes of the house, neighborhood characteristics, and other variables. Detailed information on the variables is provided in Table 1. Mobile homes, model houses, and townhouses are not considered. The sample includes 5,022 observations.

Tables 1 and Table 2 give numerical details of the distribution of TOM.[7] According to Table 1, the average time on the market for the entire sample is 56 days. According to Table 2, the mean sales duration at the 0.3 quantile is 22 days, the mean sales duration at the 0.8 quantile about 77 days. Figure 1 provides a Kaplan-Meier graph of the empirical probability of a sale relative to the time on the market in days (TOM). It shows that the probability of a house having sold by the time it is 50 days on the market is in excess of 60%. Figure 2 displays the distribution of TOM by quantiles against the background of the Rutherford county street map.[8] The darker dots represent larger TOM values. It is apparent that there is some spatial clustering of dots of the same color, which may indicate spatial correlation.

The covariates included to explain the sales duration are composed of standard variables that characterize the physical attributes of the house, location effects, such as

---

[6] The data are provided by the Middle Tennessee Regional Multiple Listing Service (MTRMLS), Realtrac Inc.
[7] Since the data is based on housing transactions, the distribution is conditional on all houses having sold.
[8] The thick red line indicates Interstate 24. The quantile map is produced in ArcGIS 9.x.

school zones, and neighborhood effects. Some common house characteristics, such as bedrooms and bathrooms, are excluded from the analysis because some preliminary screening suggested that they are collinear with square footage and have little or no additional explanatory power for TOM. The same applies to some other home features that tend to vary with square footage, such as the number of decks, patios, and garage spaces.

Since the data cover several years, an indicator variable is included for each year to capture systematic differences across time that may arise from changes in the macroeconomic environment or in the local housing market.[9] The subprime mortgage problem and the resulting downturn in real estate markets throughout the U.S. provide an example for the importance of incorporating time as a variable. Figure 3 presents a Kaplan-Meier graph of the probability of a sale against TOM for each of the years covered by the sample. The curves for the individual years reveal perceptible differences. The year 2003, which serves as the base year in all regressions, has by far the lowest sales probabilities. The probabilities of the year 2007 are close to those of 2003. 2007 is the year where the housing market crisis that followed the subprime mortgage problem began to be felt. Based on Figure 3, one would expect to see negative coefficient estimates for each of the 0/1 year indicator variables for the quantile regressions and positive coefficients for duration models of the proportional hazard type.

The models also account for seasonality, which has been shown to be a vital factor determining TOM (Haurin, 1988; Anglin, Rutherford, and Springer 2003; Kluger and Miller, 1990). More houses are typically marketed in the summer or, more generally,

[9] Adding 0/1 indicator variables is an alternative to splitting the sample, as for example in Leung et al. (2002).

in warmer weather. To account for this, a dichotomous variable is constructed to represent the winter months (*winter*). Based on the previous literature, listing in the winter season should lower the probability of a sale and, hence, raise TOM. The corresponding Kaplan-Meier graph (Figure 4) supports this idea. Figure 4 also suggests that the percentage reduction in the probability of a sale is not uniform across the distribution of TOM. This fact is an indication that quantile regression is likely to be advantageous relative to proportional hazards models in modeling TOM.

GIS data on high school zoning maps are from the Rutherford County Office of Information Technology. They allow the assignment of each house to one of the seven public high school zones in the county: Blackman, Eagleville, LaVergne, Oakland, Riverdale, Smyrna, and Siegel. Siegel High School is used as the base high school zone and, hence, not identified by a 0/1 variable. It generally has the highest ranking among Rutherford county high schools.[10] Siegel High School directly competes with Blackman, Oakland, and Riverdale High Schools in the sense that they are located in the same town. There is less competition with the other three high schools as they are located in other towns. The fact that each of the three remaining high school zones covers a separate town suggests that their coefficients will pick up general town characteristics as much as high school quality. It is expected that home sales in all high school zones that are identified with a variable in the regressions are slower than for Siegel High School because the other schools are ranked lower or are in less desirable areas of the county. Slower house

[10] The ranking is based upon historical test scores and retrieved from the Internet (www.greatschools.net). The test scores rely on the percentage of students scoring at or above the proficient level on the Tennessee Comprehensive Assessment Program (TCAP) Achievement Tests on Algebra I, Biology I, and English II. The data are compiled from the Tennessee Department of Education, 2006-2007.

sales would call for a positive coefficient in quantile regressions and a negative one in duration models.

A potentially important determinant of TOM is the price competitiveness of the house.[11] Price competitiveness (*pricedif*) is measured in this study by comparing the original listing price to the sales prices previously achieved in the neighborhood.[12] Variable *pricedif* can also be interpreted as a proxy of the seller's "eagerness" to sell. A small value indicates that the seller chose a lower price initially to sell quickly. It is expected that a house will take longer to sell the higher the original listing price is compared to the average neighborhood sales price. This calls for a positive sign in quantile regressions and a negative one in hazard models. The fact that higher list prices raise TOM is supported by Figure 5, which provides Kaplan-Meier graphs of the probability of a sale relative to TOM for expensive homes and for inexpensive ones. The probability of a sale is significantly higher for the less expensive homes. It is also apparent from Figure 5 that the decrease in the probability of selling an expensive home varies perceptively across TOM, which suggests that a quantile regression approach may be appropriate.

Unobserved neighborhood characteristics can be important in determining the duration of sales. Houses located in a desirable neighborhood or subdivision may sell faster. Figure 2 appears to provide some evidence that lightly shaded dots, which identify low values of TOM, cluster spatially. Some earlier studies have addressed the role of

[11] Many studies use functions of the selling price as determining variables of TOM, e.g. Miller (1978), Haurin (1988), Kang and Gardner (1989); Green and Vandell (1994), Forgey et al. (1996), Lee and Chang (1996), Ortalo-Magne and Merlo (2000), Huang and Palmquist (2001), and Anglin, Rutherford and Springer (2003). This approach is not taken here to avoid the estimation problems that arise from the simultaneity between TOM and selling price or between TOM and various measures of price reductions from the listing price.

[12] The concept of a neighborhood is used for the construction of a number of variables in this study. It is defined at the end of the data section.

unobserved neighborhood characteristics, but varied in how to quantify them. This paper uses a spatial lag variable for the dependent variable (*splag_lntom*). The variable contains a distance weighted average of the variable TOM for neighboring homes that sold earlier. The spatial lag is preferred over variables derived from census tract data (e.g., Zuehlke 1987) because the data are more up-to-date and focused on a smaller area.

Unusual characteristics of a house can have a major effect on TOM. Haurin (1988), Glower et al. (1998), and Capozza et al. (2005) try to capture unusualness by using the deviations of a house's observed attributes from their sample means in an indirect measurement approach. Direct measurements of atypicality would include the addition dummy variables for particular observable features, such as lakefront (Robinson and Waller, 2005), an indoor/outdoor pool (Forgey et al. 1996; Anglin et al. 2001), or for propoerties with too large or too small a lot size (Clauritie and Thistle 2007). However, one may argue that unusual characteristics that affect TOM may often be unobservable from the variables that are typically available. For example, a house may be in a "run-down" state because it was previously rented, have structural problems, may be adjacent to a commercial area and subject to noise or air pollution, or it may have an unusually expensive or tasteful interior. None of these characteristics will typically show up in MLS data. However, one can presume that unusual characteristics of this type find their way into the initial pricing of the property. If the unusual characteristics were priced according to their true market value, one may not expect to see any impact on TOM. But sellers can not typically be expected to fully account for the market discount or premium of the unusualness of their houses. A more likely scenario is that they adopt a discount that is too low and a premium that is too high, which should make the houses more difficult to sell. Two dummy variables are constructed to capture these cases. They are unity if the price per square foot is two standard deviations below or above the sample mean. One would expect that TOM rises when either variable takes on the value of unity.

A number of the variables rely on the concept of a neighborhood. The study makes use of the spatial-temporal technique of Pace et al. (2000) to identify houses in a neighborhood. To be considered as neighbors, houses need to meet two criteria. First, they need to have sold before the house of interest. This is the temporal criterion. It is crucial for the purpose of estimation because it renders the associated variables exogenous, which greatly simplifies estimation. Second, the definition of a neighborhood is meaningless if it includes houses that are located far away. This can easily happen when a nearest neighbor criterion is used to identify neighbors without a maximum allowable distance between a given house and its neighbors. The cut-off distance used in this study is 0.5 miles. Once all neighbors of a given home are identified, a standard isotropic spatial weight matrix is calculated.[13] This weight matrix is used to construct all variables that are based on neighborhood information. For example, the spatial lag variable (*splag_lntom*) is derived as the product of the weight matrix and the vector of the log values of TOM.

4 Estimation Results

The estimation results are collected in Tables 3 through 5. Each table reports on a different model. There is no change in most of the variables from model to model or in

[13] The weight is calculated by dividing the inverse of the distance in miles to neighbor *i* by the sum of the inverse distances to all neighbors.

the number of observations. For each model, five different sets of regression results are provided; four pertain to quantile regressions and one to the Cox proportional hazards model. The four quantile regressions are focused on the 0.1, 0.2, 0.5 and 0.8 quantiles of TOM. According to Table 2, these quantile points relate to TOM values of 10, 18, 42, and 86 days on the market. Both the estimated coefficients and the associated hazard ratios are provided for the Cox proportional hazards models.

A first result not apparent from the tables is that the proportionality assumptions at the heart of the Cox proportional hazards model are rejected for each model at much better than the one percent level of confidence.[14] This provides prima facie evidence in favor of a quantile regression approach. The quantile approach does not require a proportionality assumption because the coefficients are allowed to vary along the distribution of TOM. This leads to sizable coefficient differences across quantiles. For example, there is a 77% difference across the coefficients of variable *pricedif* in Table 5 and a 45% difference for variable *sqft*. Some variables are not statistically significant for low values of TOM, e.g. *psqft_hi* or *flfinwd*, but they turn significant for higher values of TOM. The reverse can also be observed, such as for variables *lotacre* or *exvinyl*. If one counts by quantile the number of covariates that are statistically significant at the 5% level or better, there are more covariates that have an impact on TOM between the second and fifth quantile, which covers roughly a sales duration between 2 ½ to 6 weeks, than at any other time. This is consistent with the Kaplan-Meier graphs of Figures 3 through 6. Despite the differences in coefficients across quantiles, the sign and size of the quantile regressions are generally consistent with those of the Cox models. One can find at least

[14] The tests of the proportional hazards assumption are based on Schoenfeld residuals. The same test results are obtained if a Weibull model is used instead of the Cox model.

one or two quantile coefficients that are close to the corresponding coefficients of the Cox model. In that sense, the quantile results do not generate true surprises. But they do provide more detailed information than the Cox models.

An important result is the absence of the spatial lag variable (*splag_lntom*) from all regressions. The variable is left out of the tables because in none of the regressions is it statistically significant at any reasonable level. This suggests that there is no detectable spatial clustering of high or low TOM values in neighborhoods. One may presume that clustering is not present in the data because sellers will reduce their initial listing price when they observe high TOM values in the neighborhood.

Models 1 through 3 differ because of the inclusion or exclusion of three variables, all of which are related in one way or another to the initial listing price. Model 2 adds to the base model (Table 3) the indicator variables *psqft_low* and *psqft_hi*, which are meant to identify whether the house is in some sense unusual. The negative coefficients of variable *psqft_low* for the quantile regressions are different from what was expected. They suggest that unusually low prices per square foot speed up rather than slow down sales. A likely explanation for the positive sign is that there are enough "bargain hunters" on the market who are willing to put up with the reasons behind a low price per square footage, such as a state of bad repair, because they specialize in fixing up and reselling houses for a profit. The signs of *psqft_hi* are positive as expected, but statistically significant only after the first few weeks that a house is on the market.

The model reported in Table 5 adds the variable *pricedif*, which is the log difference between the original listing price and the earlier selling prices of neighboring houses. Adding this variable to Model 2 (Table 4) makes the variable *psqft_low* insignificant throughout all quantiles at the five percent level or better. The coefficient values and indicators of statistical significance also decline somewhat for variable *psqft_hi*. The variable *pricedif* has a strong positive impact on TOM peaking in its influence around the median of the TOM distribution. The sign is as expected: the higher the original listing price, the longer is the sales duration.

Most of the coefficient signs of the standard variables in Tables 3 to 5 are consistent with expectations and previous findings. For example, square footage of living space (*sqft*) is positive across all quantiles and statistically highly significant. The sign suggests that larger houses take longer to sell across the board. The coefficient oscillates around 0.22 across the quantiles without too much variation. It indicates that, everything else the same, a house with an additional one thousand square feet will take about 22% longer to sell. The impact of lot size (*lotsize*) on TOM is only significant at the lower quantiles of TOM. The negative sign suggests that a large lot is a plus for newly-listed houses. It speeds up the sale between 11% and 13% per additional acre during the first six weeks on the market. The age of the house age (*age*) is negatively related to the speed of sales and significant and effectively of the same magnitude for almost all quantiles. The negative sign means that newer houses take longer to sell, a result that is consistent with the Kaplan-Meier graphs of Figure 6. The remaining three physical house attributes, wood flooring, wood or vinyl exterior, and access to city water, have little consistent impact on TOM across quantiles. The strong statistical significance found by the Cox model for three of these four variables is replicated by the quantile results only for certain parts of the distribution. Where significant, wood flooring, exterior vinyl, and city water lower TOM, while exterior wood prolongs the sales duration. The locational variables (*discenter*) and (*disnash*) do not affect TOM at any of the listed quantiles.

All indicator variables for the year sold are statistically significant for almost all quantiles. This is consistent with the Kaplan-Meier curves shown in Figure 3. The negative signs signal a better market than for 2003, the base year. Compared with the other years, year 2005 is the year with the lowest TOM values. On average, the sales duration is almost 40% lower in 2005 than in 2003. The sales duration for 2007 is close to that for the year 2003, especially for the initial weeks that a house is on the market. The seasonal variable *winter* has a statistically significant impact on the sales duration across the distribution of TOM. The positive signs for all quantiles confirm Figure 4 that houses listed in the winter will take a longer time to sell. For example, a house listed in the winter will take 22% longer in its second and third week on the market than a comparable house not listed in the winter. The negative influence is gradually declining as the house stays longer on the market and spring is approaching.

Compared with houses located in the Siegel High School zone, all other houses are harder to sell. This applies in particular to the times between the second and sixth week on the market and is more pronounced for the schools that compete with Siegel High School (Blackman, Oakland, Riverdale) in the same town. The very large coefficients for the Eagleville and LaVergne high schools suggest that the school zone is likely to proxy for the unattractiveness of the respective town.

5 Conclusion

Previous empirical analysis of time-on-market has mostly relied on duration/survival models, such as the Cox proportional hazards model. The traditional methods are restrictive because they assume covariates to have a constant effect on sales duration throughout its distribution. The paper suggests and demonstrates that quantile regression is a viable alternative to hazard models for real estate applications. It allows coefficients to vary across the distribution of the sales duration and, therefore, offers potentially more information to real estate practitioners. For numerous variables, large differences are found in the coefficients across the sales duration distribution. The empirical work is carried out on about 5,000 observations of transactions sales data for a county in Tennessee. The empirical results reject the proportionality assumptions of the duration models throughout.

The results identify numerous variables with a strong impact on sales duration. For example, the higher the original listing price is set relative to the earlier selling prices of neighboring houses, the longer it takes to sell a house. An unusually high price per square foot will have a similar effect. Larger houses tend to sell more slowly, yet older houses faster. The high school zone can have an important impact on sales duration: houses tend to sell faster in highly ranked high school zones. The study also identifies large differences in sales duration across years. The time on the market reacts perceptively to recessions and to housing market downturns, like the one in association with the subprime crisis. Seasonal effects tend to be important too: houses sell more slowly when they are initially listed in the winter.

References

Anglin, P.M., R. Rutherford, and T. Springer. (2003). The Trade-off Between the Selling Prices of Residential Properties and Time-on-the-market: The Impact of Price Setting. *Journal of Real Estate Finance and Economics,* 26, 95-111.

Asabere, P.K. and F.E. Huffman. (1993). Price Concessions, Time on the Market, and the Actual Sale Price of Homes, *Journal of Real Estate Finance and Economics,* 6, 167-174.

Asabere, P.K., F.E. Huffman, and S. Mehdian. (1993). Mispricing and Optimal Time on the Market. *Journal of Real Estate Research*, 8, 1, 149-155.

Belkin, J., D.J. Hempel, and D.W. McLeavey. (1976). An Empirical Study of Time on Market Using a Multidimensional Segmentation of Housing Markets. *AREUA Journal*, 57-75.

Capozza, D.R., R.D. Israelsen, and T.A. Thomson. (2005). Appraisal, Agency, and Atypicality: Evidence from Manufactured Homes. *Real Estate Economics*. 33, 3, 509-537.

Clauretie, T.M. and P.D. Thistle. (2007). The Effect of Time-on-Market and Location on Search Costs and Anchoring: The Case of Single-Family Properties. *Journal of Real Estate Finance and Economics*, 35, 181-196.

Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of Royal Statistical Society,* 2, 187-220

Cubin, J. (1974). Price, Quality, and Selling Time in the Housing Market. *Applied Economics,* 6, 171-187.

D'Urso, V.T., (2002). Internet Use and the Duration of Buying and Selling in the Residential Housing Market, Economic Incentives and Voting. Ph.D. thesis, MIT.

Fitzenberger B. and R. Wilke. (2005). Using Quantile Regression for Duration Analysis. *Allgemeines Statistisches Archiv*, 90, 1, 105-120.

Forgey, F.A., R.C. Rutherford, and T. Springer. (1996). Search and Liquidity in Single-Family Housing. *Real Estate Economics*, 24, 273-292.

Glower, M, D.R. Haurin, and P.H. Hendershott. (1998). Selling Time and Selling Prices: The Influence of Seller Motivation. *Real Estate Economics*, 26, 719-740.

Hayes, K.J. and L.L. Taylor. (1996). Neighborhood School Characteristics: What Signals Quality to Homebuyers? *Economic Review*, Federal Reserve Bank of Dallas, Fourth Quarter.

Haurin, D. (1988). The Duration of Marketing Time on Residential Housing. *AREUEA Journal*, 16, 4, 396-410.

Horowitz, J.L. (1992). The Role of the List Price in Housing Markets: Theory and an Econometric Model. *Journl of Applied Econometrics*, 7, 115-129.

Horowitz, J.L. and G.R. Neumann. (1987). Semiparametric Estimation of Employment Duration Models. *Econometric Reviews*, 6, 5-40.

Jud, G.D., T.G. Seaks, and D.T. Winkler. (1996). Time on the Market: The Impact of Residential Brokerage. *Journal of Real Estate Research*, 12, 447-458.

Kang, H.B. and M.J. Gardner. (1989). Selling Price and Marketing Time in the Residential Real Estate Market. *Journal of Real Estate Research*, 4, 21-35.

Kalra, R. and K.C. Chan. (1994). Censored Sample Bias, Macroeconomic Factors, and Time on Market for Residential Housing. *Journal of Real Estate Research*, 9, 253-262.

Kiefer, Nicholas M. (1988). Economic Duration Data and Hazard Functions. *Jounal of Economic Literature*, 26, 2, 646-679.

Kluger, B.D. and N.G. Miller. (1990). Measuring Residential Real Estate Liquidity. *AREUEA Journal*, 18, 2, 145-159.

Koenker, R. and G. Bassett. (1978). Regression Quantile. *Econometrica*, 46, 33-50.

Koenker, R. and K. F. Hallock. (2001). Quantile Regression. *Journal of Economic Perspective*, 15, 143-56.

Koenker, R. and O. Geling. (2001). Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. *Journal of the American Statistical Association,* 96, 454, 458-468.

Koenker, R. and Y. Bilias. (2001). Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments. *Empirical Economics*, 26, 199-220.

Koenker, R. and Z. Xiao. (2002). Inference on the Quantile Regression Process. *Econometrica*, 70, 1583-1612.

Knight, J. (2002). Listing Price, Time on Market, and Ultimate Selling Price: Causes and Effects of Listing Price Changes. *Real Estate Economics,* 30, 2, 213-237.

Miller, N.G. (1978). Time on the Market and Selling Price. *Real Estate Economics*, 6, 2, 164-174.

Miceli, T.J. (1991). The Multiple Listing Service, Commission Split, and Broker Effort. *AREUEA Journal*, 19, 548-566.

Miceli, T.J., K.A. Pancak, and C.F. Sirmans. (2000). Restructuring Agency Relationships in the Real Estate Brokerage Industry: An Economic Analysis. *Journal of Real Estate Research*, 20, 31-47.

Leung, C.K.Y., Y.C.F. Leong, and I.Y.S. Chan. (2002). TOM: Why Isn't Price Enough? *International Real Estate Review*, 5, 1, 91-115.

Pace, R. K., R. Barry, O.W. Gilley, and C.F. Sirmans. (2000). A Method of Spatial Temporal Forecasting with An Application to Real Estate Prices. *International Journal of Forecasting,* 16, 2, 229-246.

Portnoy, S. (2003). Censored Regression Quantiles. *Journal of the American Statistical Association,* 98, 464, 1001-1012.

Powell, J.L. (1984). Least Absolute Deviations Estimation for the Cencored Regression Model. *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986). Censored Regression Quantiles. *Journal of Econometrics*, 32, 143-155.

Robinson, C.D. and B.D. Waller. (2005). Time on Market: An Analysis of Lake Properties. *ASBBS E-Journal*, 1, 1.

Walden, M.L, (1990). Magnet Schools and the Differential Impact of School Quality on Residential Property Values. *Journal of Real Estate Research*, 5, 221-230

Yang, S. and A. Yavas. (1995). Bigger is Not Better: Brokerage and Time on the Market. *Journal of Real Estate Research,* 10, 23-33.

Yavas, A. (1993). Economics of Brokerage: An Overview. *Journal of Real Estate Literature*, 2, 169-195.

Yavas, A. and S. Yang. (1995). The Strategic Role of Listing Price in Marketing Real Estate: Theory and Evidence. *Real Estate Economics*, 23, 3, 347-368.

Zietz, J., E.N. Zietz, and G.S. Sirmans. (2008). Determinants of House Prices: A Quantile Regression Approach, *Journal of Real Estate Finance and Economics*, 37, 317-333.

Zuehlke, T. (1987). Duration Dependence in the Housing Market. *The Review of Economics and Statistics*, 69, 4, 701-709.

Zumpano, L.V., H.W. Elder, and E.A. Baryla. (1996). Buying a House and the Decision to Use a Real Estate Broker. *Journal of Real Estate Finance and Economics*, 13, 2, 169-181.

Table 1: Variable Definitions and Descriptive Statistics

Variable	Definition	Mean	Std. Dev.	Min	Max
Dependent Variable					
TOM	Time-on-market, in days	55.86	47.56	4	512
ln *TOM*	Log of sales duration	3.66	0.90	1.39	6.24
Continuous Variables					
splag_lntom	Weighted average values of ln *TOM* of neighborhood houses that sold before	3.79	0.44	1.73	5.89
pricdif	Logarithmic difference between original listing price and weighted average sales price of neighborhood houses that sold before	0.11	0.27	-7.28	1.90
sqft	Size of house in square feet, divided by 1,000	1.82	0.69	0.70	10.36
lotacre	Lot size in acres	0.41	0.47	0	19
age	Age of the house as of the year sold	12.34	11.65	0	107
discenter	Distance to city center, in miles	3.63	1.93	0.17	19.03
disnash	Distance to downtown Nashville, TN, in miles	27.30	5.08	14.66	39.17
Binary Variables	value = 1 if characteristic present, 0 otherwise				
psqft_low	price per square foot is two standard deviations below the average price per square foot	0.01	0.09	0	1
psqft_hi	price per square foot is two standard deviations above the average price per square foot	0.01	0.10	0	1
flfinwd	final wood flooring in house	0.42	0.49	0	1
exwood	exterior trim is made of wood	0.08	0.27	0	1
exvinyl	exterior trim is made of vinyl	0.72	0.45	0	1
water	water source is city water	0.97	0.18	0	1
winter	listed in November, December, or January	0.19	0.39	0	1
blkhigh	school district is Blackman High School	0.14	0.35	0	1
eaghigh	school district is Eagleville High School	0.00	0.02	0	1
lavhigh	school district is LaVergne High School	0.03	0.17	0	1
oakhigh	school district is Oakland High School	0.17	0.37	0	1
rivhigh	school district is Riverdale High School	0.31	0.46	0	1
smyhigh	school district is Smyrna High School	0.19	0.40	0	1
yr2004	year sold is 2004	0.14	0.34	0	1
yr2005	year sold is 2005	0.23	0.42	0	1
yr2006	year sold is 2006	0.27	0.44	0	1
yr2007	year sold is 2007	0.23	0.42	0	1

Notes: Data are from Realtrac.Inc. and the Rutherford County Office of Information Technology. The total number of observations equals 5,022.

Table 2: The Distribution of Time-On-Market by Quantile (in days)

quantile range	Observations	Mean	Std. Dev.	Min	Max
0.0 - 0.1	488	7.19	1.86	4	10
0.1 - 0.2	563	14.38	2.24	11	18
0.2 - 0.3	405	21.53	1.68	19	24
0.3 - 0.4	595	29.45	2.89	25	34
0.4 - 0.5	445	38.48	2.39	35	42
0.5 - 0.6	512	47.85	3.17	43	53
0.6 - 0.7	529	60.72	4.33	54	68
0.7 - 0.8	495	76.96	5.23	69	86
0.8 - 0.9	480	99.93	8.81	87	117
0.9 - 1.0	510	162.51	47.09	118	512

Notes: total observations equal 5,022.

Table 3: Model 1 – Estimates by Quantile Regression and Cox Proportional Hazards Model

	quantile										
	q = 0.1		q = 0.2		q = 0.5		q = 0.8		Cox PH		
variable	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	hazd rat
sqft	0.202	0.000	0.266	0.000	0.224	0.000	0.191	0.026	-0.278	0.000	0.757
lotacre	-0.105	0.006	-0.102	0.001	-0.066	0.081	0.007	0.031	0.045	0.113	1.046
age	-0.006	0.019	-0.003	0.099	-0.004	0.037	-0.003	0.002	0.002	0.208	1.002
flfinwd	-0.007	0.912	0.021	0.620	-0.070	0.072	-0.030	0.036	0.069	0.032	1.072
exwood	0.072	0.516	0.208	0.009	0.110	0.143	0.118	0.069	-0.226	0.000	0.798
exvinyl	-0.129	0.046	-0.074	0.114	0.000	0.993	-0.034	0.040	0.019	0.598	1.019
cityw	-0.232	0.093	-0.069	0.491	-0.144	0.131	-0.140	0.087	0.175	0.027	1.192
discenter	0.003	0.839	-0.013	0.219	-0.012	0.235	-0.006	0.009	0.010	0.252	1.010
disnash	-0.008	0.466	-0.001	0.899	-0.002	0.824	0.008	0.007	-0.010	0.136	0.990
blkhigh	-0.019	0.850	0.158	0.023	0.099	0.133	0.114	0.060	-0.126	0.021	0.882
eaghigh	0.636	0.032	0.054	0.862	-0.096	0.867	0.871	0.268	-0.552	0.344	0.576
lavhigh	0.097	0.602	0.189	0.170	0.157	0.227	0.225	0.116	-0.311	0.004	0.733
oakhigh	0.093	0.349	0.112	0.113	0.072	0.281	0.019	0.061	-0.031	0.583	0.970
rivhigh	0.084	0.394	0.153	0.027	0.090	0.172	0.026	0.061	-0.086	0.117	0.917
smyhigh	0.003	0.980	0.075	0.424	0.112	0.202	0.101	0.078	-0.167	0.024	0.846
yr2004	-0.099	0.315	-0.207	0.003	-0.291	0.000	-0.226	0.061	0.274	0.000	1.315
yr2005	-0.314	0.000	-0.371	0.000	-0.482	0.000	-0.372	0.054	0.495	0.000	1.640
yr2006	-0.258	0.003	-0.233	0.000	-0.345	0.000	-0.349	0.053	0.439	0.000	1.552
yr2007	-0.035	0.691	-0.047	0.456	-0.162	0.006	-0.197	0.055	0.242	0.000	1.273
winter	0.147	0.023	0.233	0.000	0.175	0.000	0.153	0.040	-0.161	0.000	0.851
constant	2.785	0.000	2.707	0.000	3.845	0.000	4.256	0.240			

Notes:coeff stands for coefficient, q = 0.1 for quantile 0.1; Cox PH for the Cox proportional hazards model and hazd rat for the hazard ratio. The estimates are based on 5,022 observations.

Table 4: Model 2 – Estimates by Quantile Regression and Cox Proportional Hazards Model

	quantile										
	q = 0.1		q = 0.2		q = 0.5		q = 0.8		Cox PH		
variable	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	hazd rat
psqft_low	-0.641	0.014	-0.490	0.011	-0.356	0.067	-0.377	0.028	0.390	0.017	1.478
psqft_hi	0.322	0.170	0.118	0.506	0.436	0.012	0.430	0.004	-0.541	0.000	0.582
sqft	0.210	0.000	0.275	0.000	0.225	0.000	0.192	0.000	-0.276	0.000	0.759
lotacre	-0.105	0.007	-0.107	0.000	-0.071	0.064	-0.035	0.226	0.057	0.035	1.059
age	-0.006	0.019	-0.003	0.124	-0.004	0.039	-0.002	0.164	0.001	0.390	1.001
flfinwd	-0.018	0.761	0.013	0.748	-0.078	0.049	-0.036	0.317	0.074	0.023	1.077
exwood	0.098	0.368	0.200	0.009	0.108	0.155	0.148	0.029	-0.233	0.000	0.792
exvinyl	-0.132	0.038	-0.078	0.082	-0.005	0.907	-0.023	0.548	0.009	0.800	1.009
cityw	-0.242	0.076	-0.078	0.423	-0.156	0.103	-0.163	0.057	0.186	0.019	1.204
discenter	0.004	0.789	-0.012	0.227	-0.013	0.234	-0.004	0.669	0.009	0.291	1.009
disnash	-0.007	0.512	-0.001	0.872	-0.001	0.920	0.007	0.324	-0.010	0.146	0.990
blkhigh	-0.021	0.828	0.184	0.006	0.112	0.090	0.110	0.064	-0.125	0.022	0.883
eaghigh	0.627	0.032	0.065	0.827	-0.079	0.891	0.843	0.001	-0.548	0.347	0.578
lavhigh	0.071	0.701	0.200	0.128	0.165	0.208	0.226	0.049	-0.312	0.004	0.732
oakhigh	0.086	0.383	0.122	0.072	0.075	0.269	0.032	0.598	-0.015	0.782	0.985
rivhigh	0.082	0.394	0.170	0.011	0.096	0.151	0.044	0.467	-0.089	0.106	0.915
smyhigh	-0.004	0.971	0.090	0.321	0.127	0.154	0.102	0.182	-0.166	0.026	0.847
yr2004	-0.131	0.176	-0.209	0.002	-0.284	0.000	-0.247	0.000	0.284	0.000	1.329
yr2005	-0.331	0.000	-0.356	0.000	-0.485	0.000	-0.401	0.000	0.505	0.000	1.657
yr2006	-0.273	0.001	-0.235	0.000	-0.348	0.000	-0.376	0.000	0.454	0.000	1.575
yr2007	-0.049	0.576	-0.048	0.432	-0.172	0.004	-0.217	0.000	0.263	0.000	1.301
winter	0.166	0.009	0.217	0.000	0.175	0.000	0.132	0.001	-0.160	0.000	0.852
constant	2.767	0.000	2.697	0.000	3.835	0.000	4.303	0.000			

Notes:coeff stands for coefficient, q = 0.1 for quantile 0.1; Cox PH for the Cox proportional hazards model and hazd rat for the hazard ratio. The estimates are based on 5,022 observations.

Table 5: Model 3 – Estimates by Quantile Regression and Cox Proportional Hazards Model

	quantile										
	q = 0.1		q = 0.2		q = 0.5		q = 0.8		Cox PH		
variable	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	hazd rat
pricedif	0.187	0.040	0.145	0.017	0.257	0.000	0.155	0.044	-0.260	0.000	0.771
psqft_low	-0.408	0.200	-0.356	0.091	-0.178	0.321	-0.310	0.085	0.200	0.253	1.222
psqft_hi	0.214	0.416	0.062	0.731	0.398	0.008	0.408	0.008	-0.467	0.001	0.627
sqft	0.176	0.000	0.248	0.000	0.181	0.000	0.167	0.000	-0.233	0.000	0.792
lotacre	-0.115	0.008	-0.130	0.000	-0.117	0.000	-0.035	0.237	0.063	0.017	1.065
age	-0.007	0.007	-0.003	0.068	-0.005	0.002	-0.003	0.082	0.002	0.283	1.002
flfinwd	-0.016	0.798	0.019	0.648	-0.089	0.009	-0.033	0.369	0.075	0.021	1.078
exwood	0.113	0.356	0.196	0.012	0.096	0.147	0.130	0.065	-0.236	0.000	0.790
exvinyl	-0.130	0.065	-0.094	0.040	0.010	0.801	-0.023	0.558	0.011	0.770	1.011
cityw	-0.245	0.108	-0.059	0.547	-0.121	0.149	-0.165	0.063	0.179	0.024	1.195
discenter	0.001	0.946	-0.013	0.218	-0.011	0.212	-0.006	0.560	0.009	0.280	1.009
disnash	-0.006	0.639	0.000	0.991	0.003	0.664	0.006	0.434	-0.010	0.137	0.990
blkhigh	0.000	0.999	0.190	0.005	0.110	0.057	0.106	0.083	-0.124	0.022	0.883
eaghigh	0.745	0.023	0.096	0.750	-0.096	0.849	0.879	0.001	-0.593	0.309	0.553
lavhigh	0.104	0.613	0.226	0.091	0.177	0.120	0.216	0.068	-0.309	0.004	0.735
oakhigh	0.093	0.400	0.119	0.086	0.041	0.484	0.019	0.760	-0.008	0.881	0.992
rivhigh	0.066	0.536	0.153	0.023	0.071	0.222	0.045	0.467	-0.083	0.131	0.920
smyhigh	-0.007	0.961	0.093	0.310	0.136	0.079	0.084	0.288	-0.151	0.043	0.860
yr2004	-0.112	0.298	-0.218	0.001	-0.265	0.000	-0.242	0.000	0.283	0.000	1.327
yr2005	-0.332	0.001	-0.389	0.000	-0.484	0.000	-0.408	0.000	0.515	0.000	1.674
yr2006	-0.298	0.002	-0.261	0.000	-0.353	0.000	-0.399	0.000	0.477	0.000	1.611
yr2007	-0.057	0.565	-0.071	0.250	-0.175	0.001	-0.244	0.000	0.287	0.000	1.332
winter	0.145	0.041	0.223	0.000	0.172	0.000	0.137	0.001	-0.165	0.000	0.848
constant	2.810	0.000	2.735	0.000	3.779	0.000	4.400	0.000			

Notes:coeff stands for coefficient, q = 0.1 for quantile 0.1; Cox PH for the Cox proportional hazards model and hazd rat for the hazard ratio. The estimates are based on 5,022 observations.

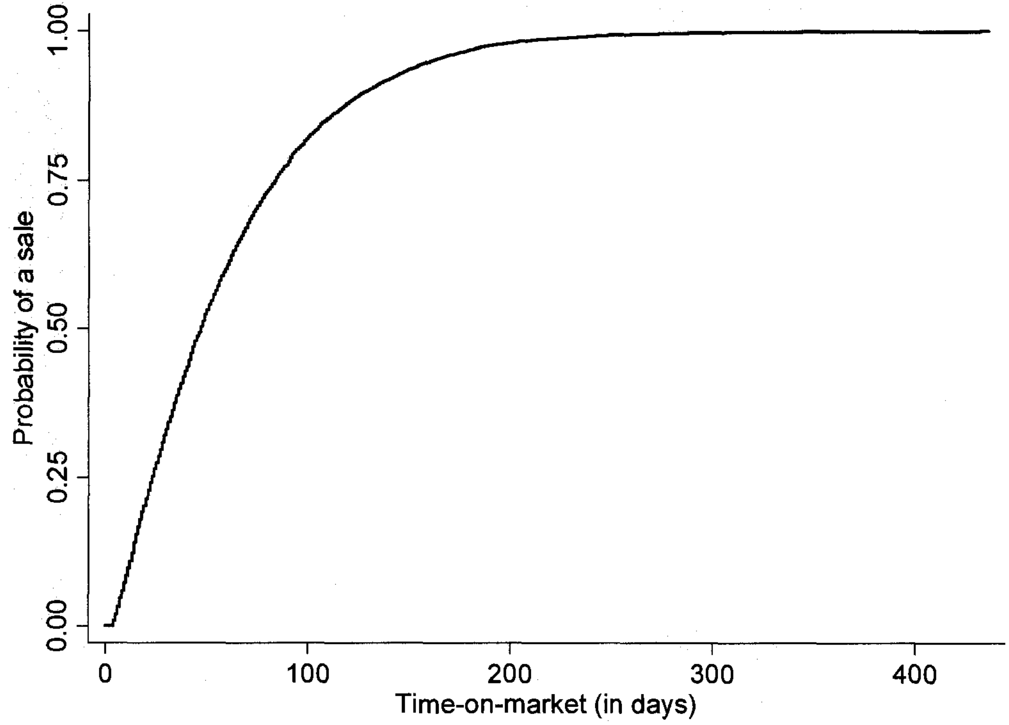


Figure 1: Probability of a Sale as a Function of Days on the Market

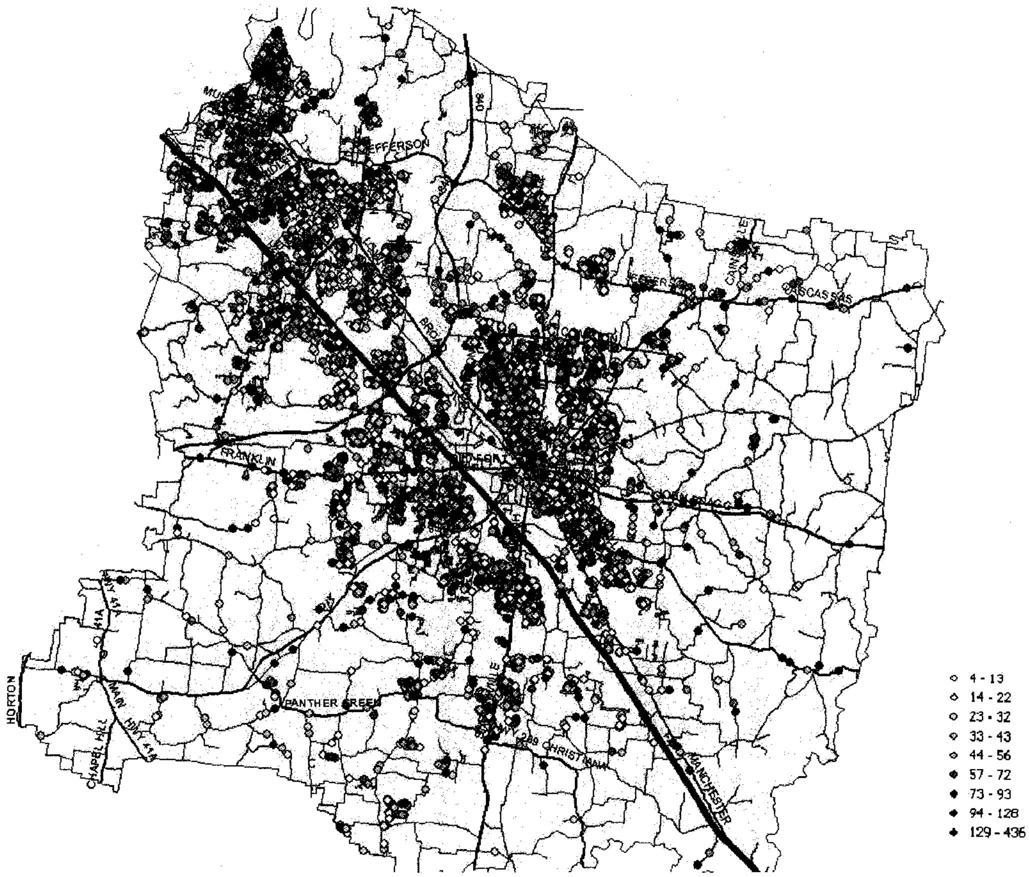


Figure 2: Distribution of TOM (in Days) in Rutherford County, TN, 2003-2007

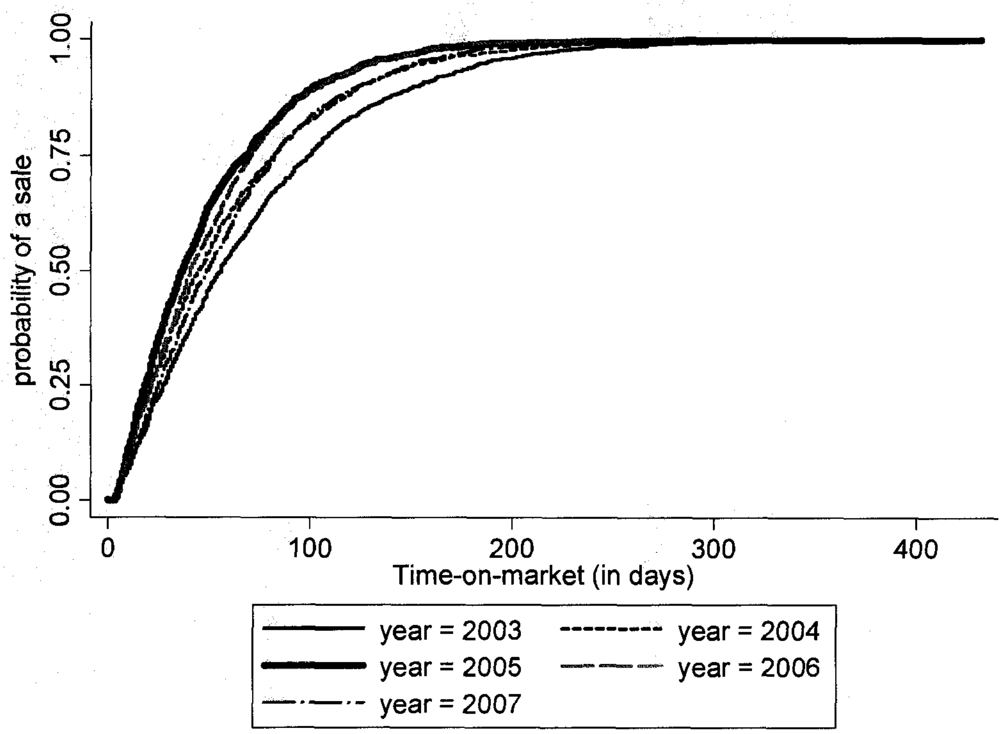


Figure 3: Probability of a Sale by Year of Sale

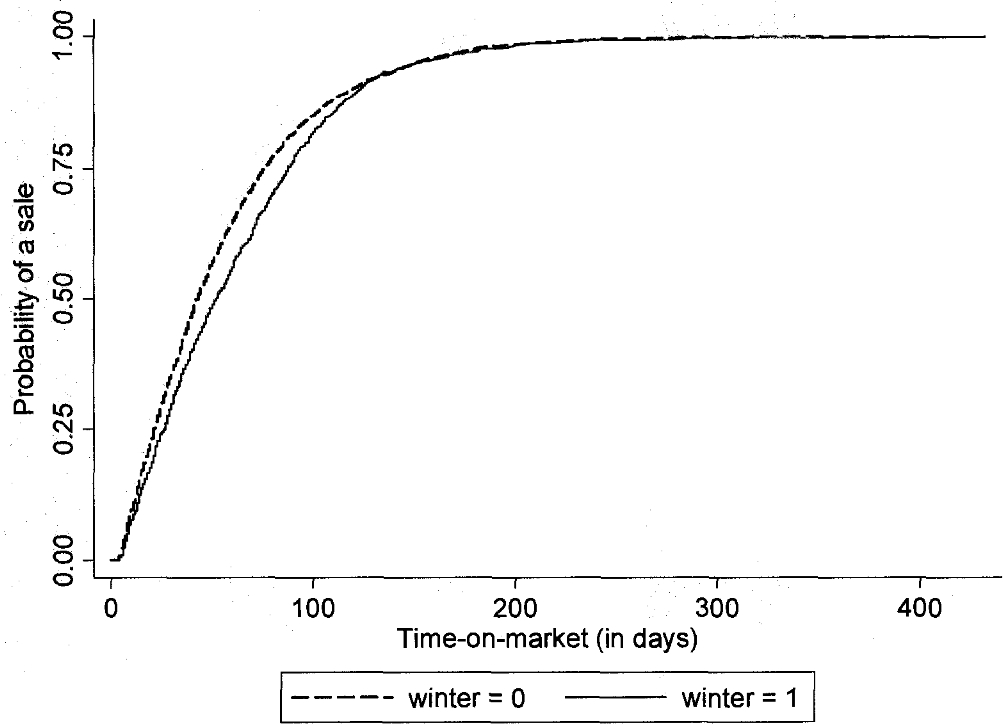


Figure 4: Probability of Selling a House That is Listed During the Winter Versus One not Listed During the Winter

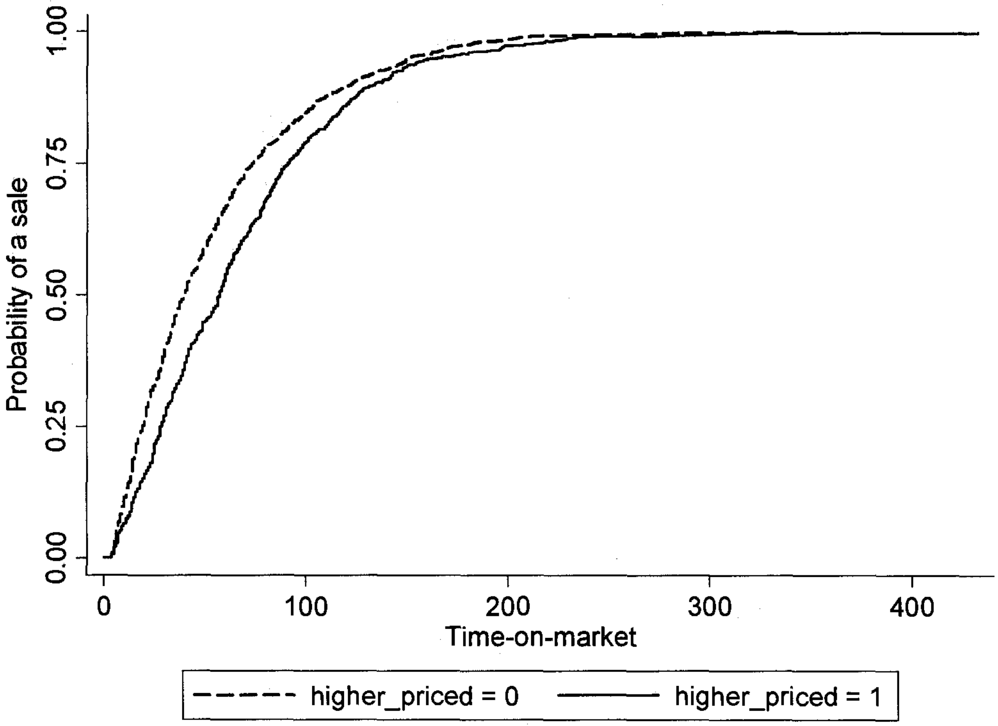


Figure 5: Probability of Selling a House in the Highest Price Quantile (> \$230,000) Versus One in the Lowest Two Quantiles ($\leq$ \$114,900)

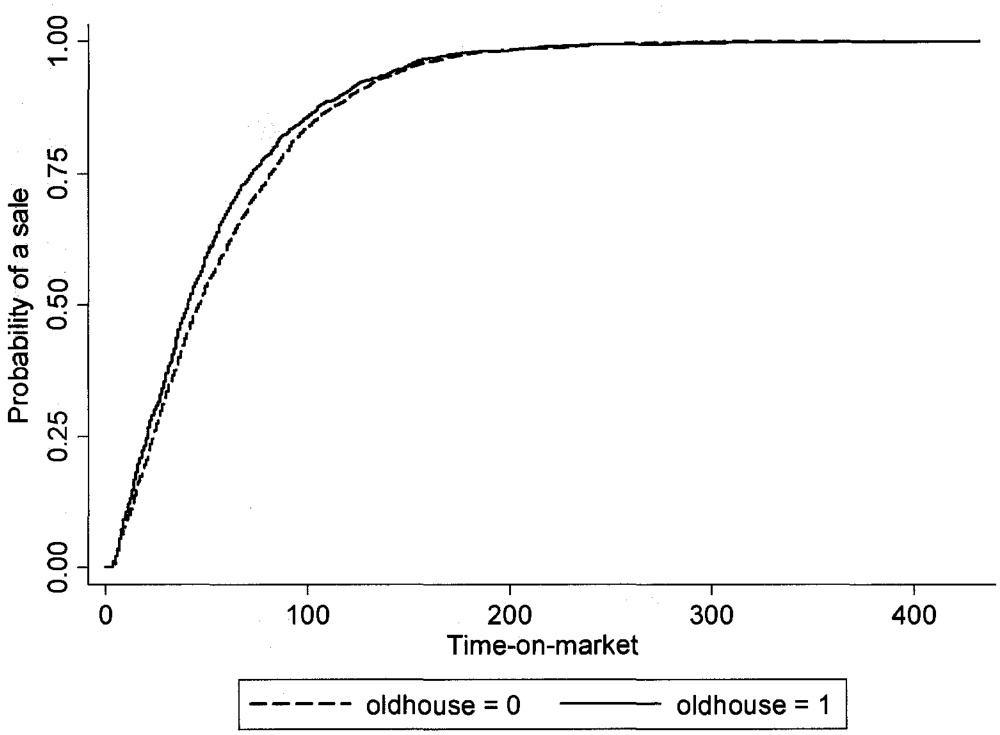


Figure 6: Probability of Selling a House in the Upper Two Age Quantiles (>16 Years) Versus One in the Lower Three Quantiles $(\leq 5\,\text{years})$