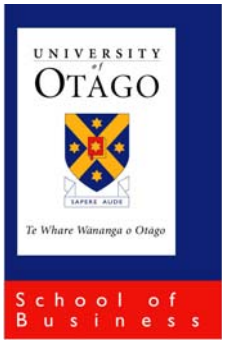

ISSN 0111-1760

University of Otago
Economics Discussion Papers
No. 0515

December 2005

Housing Renovations and the Quantile Repeat Sales Price Index

Daniel P. McMillen[1] and Paul Thorsnes

Contact details:

Paul Thorsnes
Department of Economics
University of Otago
PO Box 56
Dunedin,
New Zealand
Email: pthorsnes@business.otago.ac.nz

[1]Department of Economics (MC 144), University of Illinois at Chicago, 601 S. Morgan St., Chicago, IL 60607, mcmillen@uic.edu

Abstract

A median-based quantile estimator is less prone to the upward bias caused by unobserved renovations and missing "quality" variables when estimating appreciation rates using the repeat sales approach. As expected, the quantile estimator indicates lower appreciation rates for a sample of repeat sales of single-family homes in Chicago for 1993-2002. The quantile estimator is much less sensitive than the standard repeat sales estimator to the omission of observations for which building permits have been issued during this time. Thus, the quantile approach appears to be less prone to sample selection and missing-variable bias than the standard estimator.

1. Introduction

The purpose of a house price index is to track the rate of price appreciation over time for a standard or representative house. Using sample averages to construct the index is inappropriate because a small number of sales of high-priced homes can significantly affect a mean-based house price index. Non-academic estimates of price indexes, such as those reported by the National Association of Realtors or by local newspapers, frequently use the sample median as the basis for constructing an index. Although the median is an improvement over the mean, it does not control for house characteristics. If large, new houses dominate sales during later periods, both the mean and the median may imply an artificially high rate of price appreciation.

Academic researchers have most often used one of two mean-based methods for constructing quality-controlled price indexes. The first method is a straightforward hedonic price function, in which the natural logarithm of sales price is regressed on a vector of house characteristics and variables indicating the time of sale. The estimated coefficients on the time-of-sale variables produce the house price index. The other frequently used method is a repeat sales estimator. A repeat sales price index is estimated by regressing the percentage change in the sales prices houses that sold have more than once on a vector of discrete variables representing the time of sale. By focusing on price changes rather than levels, the repeat sales estimator avoids bias from missing house characteristics that remain unchanged over time. However, it may be subject to more severe sample selection bias than the hedonic approach because the relatively small sample of properties that sell at least twice may be even less representative of the overall population of houses. And the repeat sales model remains prone to bias from missing information on various home improvements that take place between sales.

As regression-based models, the hedonic and repeat sales approaches are mean-based procedures. As such, they are sensitive to outliers. Ironically, this sensitivity to outliers is precisely the reason that non-academic price indexes typically rely on medians rather than means to construct price indexes. In addition, regression-based procedures invoke the assumption that all estimated coefficients – including the critical time of sale variables – do not depend on whether a home sale is drawn from the tails or the middle of the house price error distribution. However, the rate of appreciation may, in fact, depend on the home's position in the error distribution. For example, recently renovated homes may appreciate especially rapidly. These observations may appear as outliers since they are likely to comprise a small portion of the overall sample. Alternatively, appreciation rates may be especially high for unusually high-quality homes or those drawn from premium locations. Variables representing renovation, high quality, and premium locations are likely to be unobserved, relegating their effects to the error term. In situations such as these, a median-based estimator may generate lower rates of appreciation than a standard, mean-based regression procedure.

In this paper, we propose the use of a quantile regression procedure to estimate repeat sales house price indexes. A median-based quantile estimator is less prone to the upward bias caused by the unobserved renovations and missing "quality" variables that are endemic to existing data sets when estimating appreciation rates. As expected, the quantile estimator indicates lower appreciation rates for a sample of repeat sales of single-family homes in Chicago for 1993-2002. Unlike most previous research, our dataset includes a variable indicating whether a building permit was issued for the home during this time. Omitting these observations leads to significantly lower appreciation rates. The quantile estimator is much less sensitive

than the standard repeat sales estimator to the omission of the observations for which building permits have been issued. Thus, the quantile approach appears to be less prone to sample selection and missing-variable bias than the standard estimator.

2. The Repeat Sales Price Indexes

The repeat sales price index can be derived from a standard hedonic price function. The hedonic approach is typified by the following equation:

$$y_{it} = \alpha + \beta' x_i + \delta_2 D_{2,it} + \ldots + \delta_T D_{T,it} + u_{it} \qquad (1)$$

In equation (1), y_{it} is the natural logarithm of the price of home i at time t, x_i is a vector of housing characteristics such as square footage and the number of bedrooms, and u_{it} is an error term. Sales dates range from 1 to T. The dummy variables $D_{2,it}$ … $D_{T,it}$ indicate that the home sold during the period represented by the first subscript. Examples of the hedonic price index approach include Kiel and Zabel (1997), Mark and Goldberg (1984), Palmquist (1980), and Thibodeau (1989).

Missing variables are probably the most important source of bias in hedonic estimates. The estimated price index will be biased if the missing variables are correlated with the time dummy variables. For example, suppose that the missing variable is a measure of house quality. If homes selling at later dates tend to be of higher quality than those from early sales, the δ's from later periods will be biased upward and will overstate the rate of price appreciation of a standard house.

The repeat sales approach was originally proposed by Bailey, Muth, and Nourse (1963). Examples include Case and Quigley (1991), Case and Shiller (1987, 1989), Follain and Calhoun (1997), and Kiel and Zabel (1997). For the subset of homes in the sample that sold at least twice, we can calculate the difference in sales

prices between time s and t, where $s < t$. The estimating equation for the standard repeat sales estimator is

$$y_{it} - y_{is} = \delta_2\left(D_{2,it} - D_{2,is}\right) + \ldots + \delta_T\left(D_{T,it} - D_{T,is}\right) + u_{it} - u_{is} \qquad (2)$$

The vector of housing characteristics, x_i, does not appear in this equation because we have assumed that the characteristics and the coefficient vector β do not change over time. If these assumptions are correct, the repeat sales estimator provides unbiased estimates of the price index without requiring data on all relevant housing characteristics. Thus, a missing variable such as house quality will not bias the estimates unless it changes over time or its coefficient changes.

The following specification accounts for missing variables and time-varying coefficients by adding a new variable, z, with values that change over time:

$$y_{it} = \alpha + \beta' x_i + \delta_2 D_{2,it} + \ldots + \delta_T D_{T,it} + z_{it} + u_{it} \qquad (3)$$

It is irrelevant whether the source of the variation in the new term is a time-varying coefficient or changes in the variable itself (as would be the case with remodeling and renovations): we can simply rewrite the model by writing z_{it} as the product of the appropriate time dummy variable and a time-varying coefficient. Equation (3) becomes:

$$y_{it} = \alpha + \beta' x_i + \delta_2 D_{2,it} + \ldots + \delta_T D_{T,it} + \left[z_i + \lambda_2 z_i D_{2,it} + \ldots + \lambda_T z_i D_{T,it} + u_{it}\right] \qquad (4)$$

and the repeat sales version of the equation is

$$y_{it} - y_{is} = \delta_2\left(D_{2,it} - D_{2,is}\right) + \ldots + \delta_T\left(D_{T,iT} - D_{T,is}\right) + \left[\lambda_2 z_i\left(D_{2,it} - D_{2,is}\right) + \ldots + \lambda_T z_i\left(D_{T,iT} - D_{T,is}\right) + u_{it} - u_{is}\right] \quad (5)$$

The new variables measure changes in z between time t and the base period. The bracketed terms in equations (4) and (5) are the error terms when z is unobserved. The missing variables are correlated with the time variables, which leads to biased estimates of the price index.

If the sample is dominated by homes that have not been renovated, the standard repeat sales approach may produce reasonably accurate estimates of the price index. Estimated indexes will be more biased the larger the proportion of homes with characteristics that have changed over time. This bias is typically referred to as a "sample selection" bias in the literature because the sub-sample of homes that have sold at least twice is generally expected to have a greater proportion of observations with changes in housing characteristics than the full sample. A median-based estimator, which places less weight on outliers than mean-based approaches, will be particularly attractive for this sub-sample since changes in housing characteristics appear as large errors when they are not observed in the dataset.

3. Quantile Regression

Mean-based procedures such as ordinary least squares are more sensitive to outliers than median-based estimators. Although outliers are occasionally simply miscoded data, at other times missing variables lead to extreme values for the error terms. An obvious example in the case of house price models is remodeling and renovations, some of which are likely to produce extremely high values for the error terms when they are not observed in the data set. The "quality" variable may also be the source of outliers: given observed housing characteristics, unusually high-quality homes will tend to have high prices and large error terms.

Unlike ordinary least squares, the target for quantile regression estimates is a parameter that is specified before estimation. Let q represent the target quantile. Also, let e_{it} be the residual implied by the econometric model. Quantile parameter estimates are the coefficients that minimize the following objective function:

$$\sum_{e_{it}>0} 2q|e_{it}| + \sum_{e_{it}\le 0} 2(1-q)|e_{it}| \qquad (6)$$

At the median, q = 0.5, which implies that equal weight is given to positive and negative residuals. At the 90th percentile, $2q$ = 1.8 and $2(1-q)$ = .2, which implies that more weight is given to positive residuals – observations with high values for the dependent variable, given the values of the explanatory variables. Equation (6) will be minimized at a set of parameter values where 100q% of the residuals are positive. This result differs from ordinary least squares, in which the only constraint on the residuals is that their *sum* equals zero.

Koenker and Bassett (1978) originally proposed the quantile regression approach. Examples of applications include Albrecht (2003); Bassett and Chen (2001); Buchinsky (1994, 1998a, 2001); Dimelis and Louri (2002); Garcia, Hernandez, and Lopez-Nicholas (2001); Hartog, Pereira, and Jose (2001); Levin

(2001); Martins and Pereira (2004); and Thorsen (1994). Buchinsky (1998b) and Koenker and Hallock (2001) present useful surveys. Each of these studies presents estimates from equations of the general form $y_i = \beta_q' x_i + u_{qi}$. The form of this equation implies that the coefficients differ by quantile. For example, Martins and Pereira (2004) find that returns to schooling are higher for more-skilled individuals. Their evidence for this conclusion comes from a regression of the natural logarithm of wages on a set of human capital characteristics, one of which is years of schooling. The coefficient for years of education is higher at higher quantiles.

Quantile effects have a straightforward missing variables interpretation that follows directly from the hedonic and repeat sales price index estimators. For example, the contribution of a sale at time t=2 to the price index can be found by taking the derivative of equation (4) or (5) with respect to $D_{2,it}$. The result, $\delta_2^* = \delta_2 + \lambda_2 z_i$, varies with the missing variable z. If $\lambda_2 > 0$, then higher values of z lead to higher values for δ_2^*. But z is part of the error term. Thus, high values of the error term imply high values for δ_2^* and low values imply low δ_2^*, i.e., quantile effects.

The intuition behind the quantile effect is the same as the motivation typically offered for selection bias in the repeat sales estimator – that the repeat sales sample is not representative of the rest of the housing market. For instance, the repeat sales sample may draw more heavily from neighborhoods with amenities that attract wealthy, mobile homebuyers, and the prices of these homes may appreciate more rapidly than homes in other neighborhoods. If the full set of neighborhood amenity variables were observed, there would be neither a quantile effect nor a sample selection issue. Similarly, recently renovated homes can be represented by a missing variable that adds to the vector of housing characteristics beginning at the time the

renovation is completed. The renovation variable produces a quantile effect because it is correlated with the time dummy variables.

The case for the quantile effect is particularly strong for the renovation example because only a minority of homes are remodeled during a given time. A renovated house shows up as an outlier in a standard regression model. Such outliers are drawn from the upper tails of the error distribution. The effects of this unobserved variable would not contaminate other points in the distribution. A median-based estimate (q = .5) will be far less vulnerable than a mean-based estimate to the effects of omitted variables that affect only a portion of the sample.

4. A Monte Carlo Analysis

In this section, we report the results of a set of Monte Carlo experiments that illustrate the benefits of the quantile approach to estimating house price indexes. The basis for the experiments is a straightforward two-period version of equation (4):

$$y_i = 5 + x_i + .2D_i + \lambda z_i D_i + u_i \tag{7}$$

The time subscript is suppressed from equation (7) because it unnecessarily complicates the notation of this simple hedonic model, which is sufficient for illustrating the benefits of the quantile approach. We draw values of x from a unit normal distribution. We generate the time variable D, by making draws from a $U(0,1)$ distribution and setting $D = 1$ when the randomly drawn value is greater than 0.5. The "missing" variable, z, is drawn from a $U(-.5,.5)$ distribution. Finally, we draw values for the error term, u, from a normal distribution with a mean of zero and a variance that assures that the R^2 from a regression of y on x, D, and zD will be approximately 0.9 on average. We let the values of λ vary from 0 to 1 while maintaining each of the

other parameters at the values shown in equation (7). Thus, observations with higher values of z have higher appreciation rates on average. Each experiment has 1000 observations.

When a sale occurs during the base time period, $D = 0$. The price of an identical home is $0.2+\lambda z$ higher when the sale takes place during the second period. If z represents quality, then the appreciation rate is higher for high-quality homes. This variable would be missing in a typical house-price study. If z is not observed, appreciation rates are higher for observations drawn from the upper tails of the error distribution. Thus, the Monte Carlo setup generates quantile effects, in which the implied marginal effect of D, i.e., $0.2+\lambda z$, varies across the error distribution.

Table 1 reports means and standard deviations for quantile regression estimates of 1000 replications of each experiment. We estimate each regression at target quantiles of 0.25, 0.50, and 0.75. The explanatory variables for the regressions are simply x and D; zD is not included. The missing variable, zD, is not correlated with x but it is correlated with D. Thus, omitting zD does not bias the estimated coefficient for x but does lead to biased estimates for the D coefficient. The true coefficient for D rises with λ, and when $\lambda>0$ it is higher at higher quantiles. Therefore, the question in the Monte Carlo analysis is whether the quantile approach indicates higher appreciation rates – i.e., higher coefficients for D – at higher quantiles. Given the structure of the Monte Carlo setup, the true coefficient is 1.0 for x at all values of λ. The true intercept is lower at lower regression quantiles because errors are negative on average at $q = 0.25$ and positive at $q = 0.75$. All calculations are performed using the QREG command in STATA.

The results are precisely as expected. The average estimated coefficient for x is close to 1.0 across the three target quantiles and across the five alternative values of

λ. Since the error term, u, and the omitted variable, λzD, both have means of zero, the estimated intercepts are approximately equal to their true value of 5.0 at the residual median ($q = 0.5$). The average intercepts are lower than 5.0 at $q = 0.25$ and are higher than 5.0 at $q = 0.75$. As expect, estimated appreciation rates – the coefficient for D – are approximately equal to the correct value of 0.20 when quantile effects are absent ($\lambda = 0$). Importantly, estimated appreciation rates are lower than 0.20 when $\lambda>0$ and $q = 0.25$, and they are higher than 0.20 when $\lambda>0$ and $q = 0.75$. The estimated appreciation rates average just under 0.20 at the median of the distribution of residuals.

The last three rows of the table show the percentage of rejections for the null hypothesis of equal coefficients for the 25% and 75% quantiles. The tests are based on 20 replications of a bootstrap algorithm. As quantile effects are absent for x, we should expect the null hypothesis to be rejected no more than 5% of the time (the nominal size of the test) for this variable. Rejection rates are somewhat lower than 5% for this variable, and they do not vary systematically by λ. Since quantile effects *always* exist for the intercept, the tests always reject the null hypothesis of equal intercepts at the 25% and 75% quantiles. The most important finding is that the rejection rate for equal coefficients for D rises with λ. This result means that, as expected, the statistical test is more likely to indicate quantile effects as the magnitude of the missing variable (λzD) increases.

To put these results in perspective, assume that z represents a trait such as renovation or simply the change in quality between the two periods. Prices of homes with positive values for z increase relatively quickly over time, and prices rise more slowly (or even fall) when z is negative. If z is unobserved, standard estimates will typically be biased. The bias is upward if quality improved in most of the homes for

which quality changed over the time period. The quantile estimator can detect differences in appreciation rates by allowing for differences in coefficients across target quantiles. In a conventional case of remodeling, most values of z equal zero while a small percentage are positive. Standard appreciation rate estimates will again be biased upward in this case. In contrast, a median-based estimator will provide accurate estimates, and the estimates at high target quantiles will detect the higher rates of appreciation associated with remodels.

5. Data and Model Specification

The dataset for the empirical application of the quantile regression estimator was drawn from sales data gathered by the Illinois Department of Revenue (IDOR). IDOR conducts reviews of assessment practices for all counties in Illinois, including Cook County. Through a Freedom of Information Act request, IDOR provided data on all sales of single-family homes in the City of Chicago for 1993-2002. The key variables are the sales price, date of sale, and the parcel identification number (or "PIN"). The PIN allows us to identify repeat sales. There were 89,806 sales during this period, of which 24,533 were sales of homes that sold more than once. The sub-sample of 24,533 sales is composed of 12,792 repeat sales pairs.

Figure 1 shows the average and median sales prices for the full sample of sales. Figure 2 shows the comparable data for the sample of homes that sold more than once. For the full sample, average prices rose from $116,790 in 1993 to $224,488 in 2002, an increase of 92.2%, or 7.5% per year on average. The average prices of repeat sales homes increased more rapidly – from $122,293 in 1993 to $254,869 in 2003, an increase of 108.4%, or 8.5% annually. The distribution of sales prices is dominated by lower-priced homes. In the full sample, the median sales price

was $106,888 in 1993 and $183,000 in 2002, an increase of 71.3% or 6.2% annually. For the repeat sales sample, the median sales price was $110,000 in 1993 and $206,000 in 2002, an increase of 87.3% or 8.2% annually.

One reason for the large difference between the appreciation rates for the full sample and the sample of repeat sales may be a high probability that homes that sell more than once are remodeled or renovated between sales. Building permit data allow us to identify many of the homes that have been altered significantly between sales. Primary permit categories include additions, alterations, new construction, and repairs. Building permits were issued between sales for 1,372 of the 12,792 repeat sales pairs in our sample (10.7%). Average sales prices for the repeat sales for which building permits were issued rose from $126,410 in 1993 to $343,119 in 2002, an increase of 171.4% or 11.7% annually. Median sales prices for these homes rose from $110,000 in 1993 to $249,000 in 2002, an increase of 126.4% or 9.5% annually. Thus, homes with building permits tend to have prices similar to other homes in 1993, but their appreciation rates are much higher on average. The higher appreciation rate is at least in part due to the failure to hold quality constant: building permits lead to changes in housing characteristics that improve the quality of the home.

Our data set includes sales from 10 years, or 40 quarters. As shown in equation (8), the standard repeat sales estimator includes 39 explanatory variables once the variable representing the first quarter is omitted to impose that the intercept equals zero. A positive value for the intercept would imply an increase in prices even within a single time period. Although within-period price increases are possible, most authors impose that the price index equals zero during the base period. (An exception is Goetzmann and Siegel (1995), who suggest including an intercept because properties are often upgraded around the time of a transaction.) Imposing a zero

intercept is not as simple for the quantile estimator. The quantile estimator conditions on points in the error distribution. At the 90th percentile, for example, the intercept will clearly be positive because the errors are positive on average when the estimator focuses on the 90th percentile of the error distribution. In our empirical application, we focus on the median, so the assumption of a zero intercept is reasonable.

In a standard repeat sales model, we can impose that the intercept equals zero in two ways. The obvious one – omitting the constant term from the regression – is not an option in quantile regression because intercepts cannot equal zero across all quantiles. The second alternative is to estimate the regression with an intercept, and then solve for the restricted least squares estimates that are implied by a zero intercept. Let X be the matrix of explanatory variables for the unrestricted regression, and let R be a vector with a one in the position corresponding to the intercept in X and zeros elsewhere. The formula for the restricted coefficients is $\hat{\beta}_r = \hat{\beta} + (X'X)^{-1} R'\left[R(X'X)^{-1} R'\right]^{-1} R\hat{\beta}$. Let s_{ij} represent the entry in row i and column j of the estimated covariance matrix, and assume that a vector of one's is the first column of X. Then the formula for $\hat{\beta}_{ri}$ – the coefficient in row i of $\hat{\beta}_{ri}$ – is $\hat{\beta}_i - \hat{\beta}_1 (s_{1i} / s_{11})$. Calculating the restricted price index by imposing a zero intercept is not equivalent to obtaining an unrestricted estimate with a non-zero value in the base period and subtracting the intercept from all dates – a parallel shift in the price index. The formula for restricted coefficients *rotates* the price index so that the restricted intercept is zero. This transformation also is a logical basis for the quantile repeat sales estimates when the target point is the median of the error distribution. This transformation allows us to directly compare the results to the standard repeat sales estimator.

6. Estimated Price Indexes

Figure 3 shows the estimated price indexes and Table 2 shows the value of the indexes for the fourth quarter of each year in the sample. Using the full sample of sales, the standard repeat sales estimator indicates that prices rose by 77.8% from the beginning of 1993 to the fourth quarter of 2002. In contrast, the median-based quantile estimator indicates that prices rose by only 68.9% over this time. In light of the previous discussion, the difference in results can be explained in at least two, complementary ways. First, there clearly are quality improvements in a significant portion of the sample since building permits were issued between sales for 10.7% of the observations. Second, there may be missing variables that do not change over time and that are correlated with the appreciation rate. For example, higher quality homes or homes in certain neighborhoods may appreciate faster than others. If the effect of time-invariant missing variables is confined to a relatively small portion of the sample, then the median-based estimator will be less sensitive to these observations than the standard estimator.

Though we cannot account for the effects of unknown missing variables, the building permit data allow us to measure the effect of renovations on the estimated price indexes. We re-estimate the models after omitting the 1,372 repeat sales pairs for which a building permit was issued between sales. The results are again shown in Figure 3 and Table 2. As expected, both the standard and the quantile repeat sales price indexes are lower after the observations with building permits have been omitted. The standard repeat sales price index is 0.734 in the fourth quarter of 2002 after the building permit observations are eliminated, compared with 0.778 when these observations are included. Comparable figures for the quantile estimator are 0.673 and 0.689. Importantly, omitting the permit data has less of an effect on the

quantile estimates than on the standard repeat sales estimates. The difference between the fourth-quarter values of the price indexes are 0.044 for the standard estimator compared with only 0.016 for the quantile estimator. Even though the quantile estimator focuses on the median, the building permit observations have some effect on the estimates because all observations receive some weight in the estimation procedure. However, the effects of the building permit observations are much less pronounced in the median based estimator.

The term "sample selection" typically is used in the repeat sales literature to refer to the percentage of the sample that fails to meet the stringent assumptions of the estimator. When the building permit data are included, we know that at least 10.7% of the observations fail to meet the condition that housing characteristics do not change over time. Table 3 shows what happens when this percentage is increased to 33.33%. We keep the 1,372 repeat sales pairs for which building permits were issued between sales, and randomly draw 2,744 pairs from the remaining, non-permit observations. The estimated rate of price appreciation increases for both the standard and the quantile repeat sales price estimator. The fourth-quarter value of the price index is 0.843 for the standard estimator compared with 0.778 when all non-permit data are included, while comparable values for the quantile estimator are 0.689 and 0.715. These results imply that the quantile estimator is much less sensitive to the "bad" data than the standard estimator.

7. Conclusion

The quantile approach has advantages over conventional mean-based approaches to estimating house price indexes. Targeting quantiles from the middle of the error distribution reduces the effects of outliers. The problem of outliers is

particularly important for the repeat sales estimator, which is vulnerable to an upward bias when the sample includes renovated houses and there is no way to identify which homes have been upgraded. In this situation, a more realistic view of the housing market may be gained by constructing indexes using lower quantiles as the target point.

Data for Chicago from 1993-2002 illustrate this advantage of the quantile approach. Our data set is unique in that we were able to merge building permit data with the sample of repeat sales in order to identify homes that were modified between sales. As expected when a substantial portion of the homes in the sample are being upgraded, the standard estimator appears to overstate the rate of appreciation in the full sample. When the sample is restricted to repeat sales pairs without building permits, estimated appreciation rates fall for either estimator. Importantly, the inclusion of building-permit data has much less of an effect on the quantile estimator; the quantile estimates will contain less upward bias in the common case in which renovations are unobserved.

The quantile approach is consistent with the standard non-academic practice of working with medians rather than averages in order to reduce the effect of outliers. In the context of our dataset, one set of outliers includes homes for which building permits have been issued. More generally, outliers include any omitted variable that has a significant effect on a subset of the observations. By reducing the influence of outlier observations, the quantile approach reduces the sample selection problems that have plagued standard applications of the repeat sales estimator.

References

Albrecht, James, Anders Bjorklund, and Susan Vroman, "Is There a Glass Ceiling in Sweden?," *Journal of Labor Economics* 21 (2003), 145-177.

Bailey, M.J., R.F. Muth, and H.O. Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association* 58 (1963) 933-942.

Bassett, Gilbert W., Jr., and Hsiu-Lang Chen, "Portfolio Style: Return-Based Attribution using Quantile Regression," *Empirical Economics* 26 (2001), 293-305.

Buchinsky, Moshe, "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62 (1994), 405-58.

Buchinsky, Moshe, "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach," *Journal of Applied Econometrics* 13 (1998a), 1-30.

Buchinsky, Moshe, "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources* 33 (1998b), 88-126.

Buchinsky, Moshe, "Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S.," *Empirical Economics* 26 (2001), 87-113.

Case, Bradford and John M. Quigley, "The Dynamics of Real Estate Prices," *Review of Economics and Statistics* 73 (1991), 50-58.

Case, Karl E. and Robert J. Shiller, "Prices of Single-Family Homes since 1970: New Indexes for Four Cities," *New England Economic Review* (1987), 45-56.

Case, Karl E. and Robert J. Shiller, "The Efficiency of the Market for Single-Family Homes," *American Economic Review* 79 (1989), 125-137.

Dimelis, Sophia and Helen Louri, "Foreign Ownership and Production Efficiency: A Quantile Regression Analysis," *Oxford Economic Papers* 54 (2002), 449-469.

Follain, James R. and Charles A. Calhoun, "Constructing Indices of the Price of Multifamily Properties using the 1991 Residential Finance Survey," *Journal of Real Estate Finance and Economics* 14 (1997), 235-255.

Garcia, Jaume, Pedro J. Hernandez, and Angel Lopez-Nicolas, "How Wide is the Gap?

An Investigation of Gender Wage Differences Using Quantile Regression," *Empirical Economics* 26 (2001), 149-167.

Goetzmann, William N. and Matthew Spiegel, "Non-Temporal Components of Residential Real Estate Appreciation," *Review of Economics and Statistics* 77 (1995), 199-206.

Hartog, Joop, Pedro T. Pereira, and A. C. Jose, "Changing Returns to Education in Portugal during the 1980s and Early 1990s: OLS and Quantile Regression Estimators," *Applied Economics* 33 (2001), 1027-1037.

Kiel, Katherine A. and Jeffrey E. Zabel, "Evaluating the Usefulness of the American Housing Survey for Creating Housing Price Indices," *Journal of Real Estate Finance and Economics* 14 (1997), 189-202.

Koenker, Roger and Gilbert W. Bassett, Jr., "Regression Quantiles," *Econometrica* 46 (1978), 33-50.

Koenker, Roger and Kevin F. Hallock, "*Quantile Regression*," *Journal of Economic Perspectives* 15 (2001), 143-156.

Levin, Jesse, "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement," *Empirical Economics* 26 (2001), 221-246.

Mark, J.H. and M.A. Goldberg, "Alternative Housing Price Indices: An Evaluation," *AREUEA Journal* 12 (1984), 30-49.

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

McMillen, Daniel P., "The Return of Centralization to Chicago: Using Repeat Sales to Identify Changes in House Price Distance Gradients," *Regional Science and Urban Economics* 33 (2003), 287-304.

Palmquist, R.B., "Alternative Techniques for Developing Real Estate Price Indexes," *Review of Economics and Statistics* 66 (1980), 394-404.

Thibodeau, Thomas G., "Housing Price Indexes from the 1973-83 SMSA Annual Housing Survey," *AREUEA Journal* 17 (1989), 110-117.

Thorsen, James A., "The Use of Least Median of Squares in the Estimation of Land Value Equations," *Journal of Real Estate Finance and Economics* 8 (1994), 183-190.

Table 1
Monte Carlo Results

Variable, Percentile	$\lambda = 0$	$\lambda = .25$	$\lambda = .50$	$\lambda = .75$	$\lambda = 1$
x, 25%	1.001 (0.015)	1.003 (0.015)	1.006 (0.016)	1.009 (0.015)	1.011 (0.016)
x, 50%	1.000 (0.014)	1.003 (0.014)	1.006 (0.014)	1.009 (0.014)	1.012 (0.015)
x, 75%	1.000 (0.015)	1.003 (0.014)	1.007 (0.016)	1.010 (0.016)	1.013 (0.017)
D, 25%	0.199 (0.029)	0.194 (0.030)	0.180 (0.030)	0.153 (0.030)	0.119 (0.031)
D, 50%	0.199 (0.027)	0.199 (0.026)	0.199 (0.027)	0.198 (0.028)	0.198 (0.030)
D, 75%	0.199 (0.028)	0.204 (0.028)	0.220 (0.030)	0.244 (0.030)	0.278 (0.032)
Intercept, 25%	4.772 (0.020)	4.771 (0.022)	4.770 (0.021)	4.768 (0.021)	4.765 (0.021)
Intercept, 50%	5.000 (0.019)	5.000 (0.019)	5.000 (0.018)	5.001 (0.019)	5.001 (0.020)
Intercept, 75%	5.228 (0.020)	5.229 (0.021)	5.231 (0.021)	5.233 (0.021)	5.236 (0.021)
Rejections of Equal Coefficients for *x* at 25% and 75%	3.0%	2.3%	3.9%	2.7%	3.5%
Rejections of Equal Coefficients for *D* at 25% and 75%	3.1%	5.1%	19.4%	65.7%	96.3%
Rejections of Equal Coefficients for Intercepts at 25% and 75%	100%	100%	100%	100%	100%

Note. Means and standard deviations (in parentheses) are reported for 1000 simulations. The base model is $y = 5 + x + .2D + \lambda zD + u$, where $z \sim U(-.5,.5)$.

Table 2
Estimated Fourth Quarter Price Indexes

Year	Standard, All Sales	Standard, No Permits	Quantile, All Sales	Quantile, No Permits
1993	0.007	0.017	0.033	0.036
1994	0.035	0.051	0.049	0.051
1995	0.067	0.077	0.096	0.095
1996	0.134	0.140	0.130	0.135
1997	0.198	0.188	0.178	0.175
1998	0.318	0.293	0.256	0.242
1999	0.433	0.406	0.345	0.335
2000	0.550	0.508	0.492	0.478
2001	0.676	0.628	0.578	0.552
2002	0.778	0.734	0.689	0.673

Table 3
Estimated Fourth Quarter Price Indexes:
High Percentage of Permit Sales

Year	Standard, All Sales	Standard, No Permits	Quantile, All Sales	Quantile, No Permits
1993	-0.069	-0.051	-0.003	0.025
1994	-0.016	0.026	0.026	0.039
1995	-0.002	0.021	0.038	0.053
1996	0.109	0.143	0.105	0.127
1997	0.217	0.201	0.155	0.157
1998	0.364	0.297	0.254	0.236
1999	0.448	0.361	0.360	0.320
2000	0.600	0.481	0.523	0.480
2001	0.760	0.641	0.619	0.543
2002	0.843	0.723	0.715	0.658

Figure 1
Sales Price Trends for the Full Sample of Sales

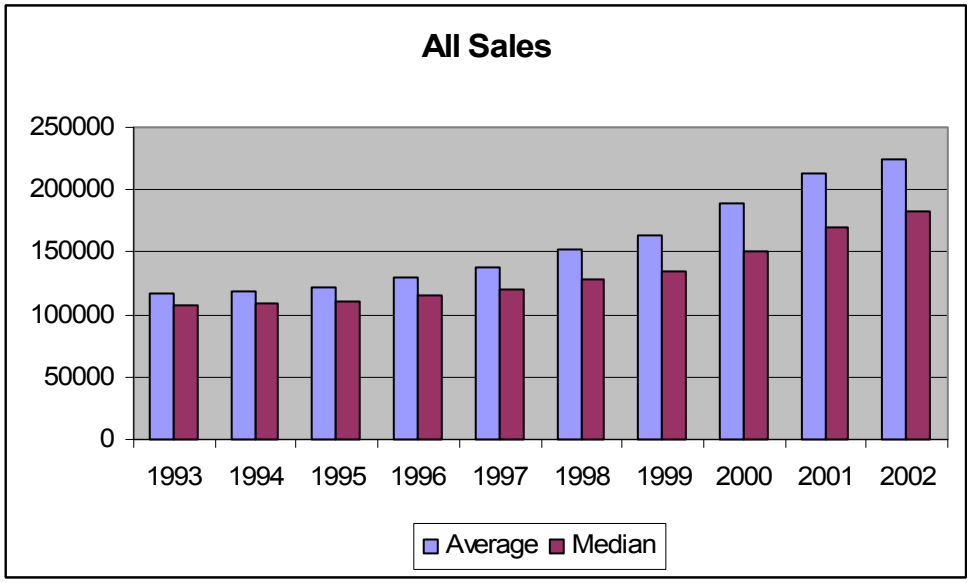

Figure 2
Sales Price Trends for the Repeat Sales Sample

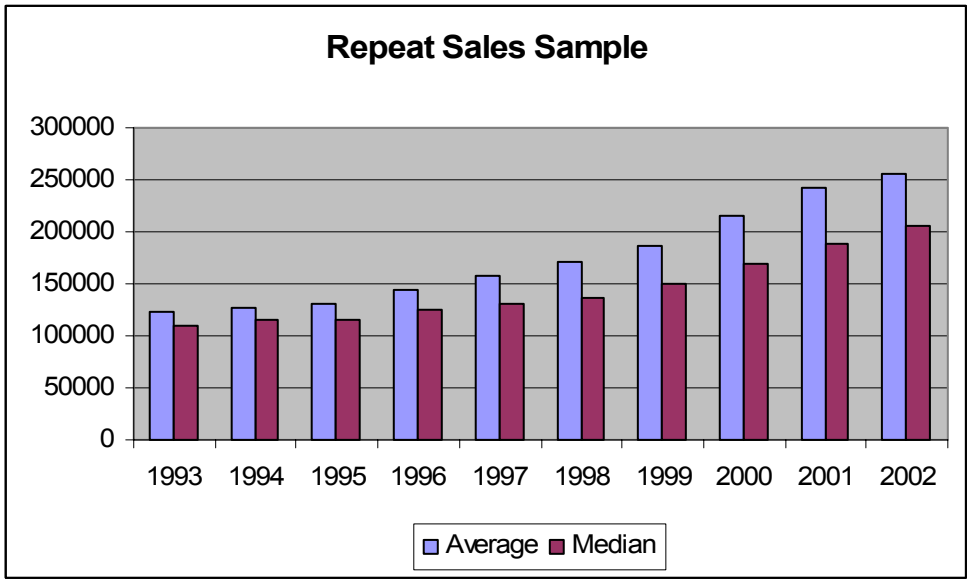

Figure 3
Estimated Price Indexes