Changes in the Distribution of House Prices over Time:

Structural Characteristics, Neighborhood, or Coefficients?

Daniel P. McMillen
Department of Economics and
Institute of Government and Public Affairs
University of Illinois at Chicago (MC 191)
815 W. Van Buren St., Suite 525
Chicago, IL 60607
mcmillen@uic.edu

November 4, 2007

Abstract

Quantile hedonic house price function estimates imply that appreciation rates were higher between 1995 and 2005 for high-priced homes in Chicago. Decompositions of temporal changes in the house price distribution suggest that the types of homes sold and their location do not account for the change in the price distribution. Rather, higher appreciation rates for high-priced homes are explained by differences in the quantile regression coefficients over time.

1. Introduction

A standard hedonic or repeat sales price index provides a single measure of the quality-adjusted price of housing at a given date. The interpretation of an increase in the price index from 1 in the base period to 1.1 in a subsequent period is that the price of a representative home increased by 10%. The implication is that *all* homes share in this price increase, or at least that any departure from the base estimate is randomly distributed around the estimated value. In practice, participants in the housing market recognize that appreciation is not shared equally by all homes. Luxury homes may continue to appreciate rapidly in slow markets, for example; or entry-level homes may appreciate especially rapidly in tight market conditions.

This study differs from most previous work on housing in its focus on changes in the full *distribution* of prices. Using data on sales of single-family homes in Chicago, I find that the distribution of house prices became less skewed between 1995 and 2005. Though prices increased throughout the distribution between 1995 and 2005, the rate of appreciation was particularly rapid for homes with higher prices, leading to a thicker distribution on the right. The obvious next question is what caused this change in the distribution? Did the distribution change simply because homes sold more frequently in high-priced neighborhoods in 2005 than in 1995, or because the 1995 sample of sales had fewer large, high-quality homes on big lots? Alternatively, is the distributional change unrelated to the size, quality, and location of the homes in sample, instead being caused by changes in the underlying hedonic price functions?

To address these questions, I use an approach developed by Machado and Mata (2005) to decompose changes in the distribution of house prices into the portion induced by changes in the distribution of the explanatory variables and the portion caused by changes in the coefficients of quantile regression estimates of the hedonic price function. I further decompose the changes into the portion due to changes in the location of the sales and changes in the structural characteristics represented in the sample. The quantile approach is ideally suited to this analysis because it directly measures the effect of an explanatory variable on different target points of the overall house price distribution. The results indicate that nearly the entire change in the distribution of house prices can be explained by changes in the coefficients of the hedonic price functions rather than by changes in the distributions of the explanatory variables or the location of the sales. Changes in the explanatory variables – particularly building area – have more effect on the distributional change than the location of the sales. Thus, the results suggest that sample composition is *not* the source of the change in the house price distribution. Rather, bigger homes in higher-priced neighborhood homes simply appreciated more rapidly than other homes between 1995 and 2005; i.e., the coefficients of the hedonic price functions have changed in such a way to increase the return on high-priced homes.

The analysis is closely related to the literature on changes in earnings inequality. Many researchers have noted that earnings have risen more rapidly for high-income workers, causing the distribution of earnings to have become more unequal over time (e.g., Buchinsky (1994, 1998a); Dickey (2007); Gosling, Machin, and Meghir (2000); Juhn, Murphy, and Pierce (1993); Katz and Murphy (1992); Martins and Pereira (2004)). A change in the return to education – a change in the coefficients of the hedonic wage function – appears to explain much more of the change in the distribution of earnings than any change in levels of schooling. If earnings have risen rapidly for high-wage workers, it perhaps is not surprising that high-priced homes have appreciated more rapidly than homes in the lower end of the house price distribution. However, since the starting house price distribution was more highly skewed than the earnings distribution and both ends of the distribution shared in the general appreciation of house prices, the 2005 distribution implies a *reduction* in the degree of house price inequality. The difference between the change in the house price distribution and the earnings distribution is that low-priced homes shared in the general rate of appreciation; high-priced homes simply appreciated more.

2. Empirical Approach

For linear regression models, Oaxaca's (1973) decomposition is commonly used to show how the conditional mean responds to changes over time in the explanatory variables and the estimated coefficients. Following conventional practice, assume that the natural logarithm of the sales price of a home, y, is a simple function of a set of structural characteristics, X, and a set of neighborhood dummy variables, D. Thus, $y = \alpha + X\beta + D\gamma + u = Z\lambda + u$, where u is an error term, $Z = \begin{pmatrix}1 & X & D\end{pmatrix}$ is the matrix of explanatory variables, and $\lambda = \begin{pmatrix}\alpha & \beta & \gamma\end{pmatrix}'$ is the coefficient vector. Using subscripts to denote time, the Oaxaca decomposition of the change in the conditional expectation of house prices between period 0 and period 1 is:

$$E(y_1 - y_0) = (Z_1 - Z_0)\lambda_1 + Z_0(\lambda_1 - \lambda_0) \quad (1)$$

The first term on the right hand side shows the effect of changes in the values of the explanatory variables on the conditional expectation, while the second term shows the effect coefficient changes. Although equation (1) is typically calculated at mean values of Z_0 and Z_1, it can just as easily be evaluated at any given set of values for the explanatory variables. Whether evaluated at the mean, median, or some other value of Z, equation (1) is based on the underlying conditional expectations $E(y \mid Z)$, which imply a single expected value of y for each value of Z.

Basing their procedure on a quantile regression estimator, Machado and Mata's (2005) approach is more general than the conventional Oaxaca decomposition. The coefficients of quantile regressions vary across quantile, q:

$$y = Z\lambda_q + u_q \tag{2}$$

Thus, the marginal effect of Z at the median is $\lambda_{0.5}$ while the marginal effect at the 90th percentile is $\lambda_{0.9}$. Since λ varies across quantiles, equation (2) implies a distribution of values for y for each value of Z. To simulate the distribution of y for any value of Z, we could first estimate λ_q for different quantiles (e.g., λ_q = 0.01, 0.02, …, 0.99) . Next, we could draw randomly from the (uniformly) distributed values of $\hat{\lambda}_q$ and calculate the implied distribution of values of y at the qth quantile. After repeating this bootstrap-style procedure B times, an estimate of the density function f(y|Z) can be obtained using standard procedures, such as a kernel density estimator or a simple histogram.

Machado and Mata (2005) carry this simulation exercise a step further by also simulating the distribution of the covariates. The steps are:

1. Estimate quantile regressions for Q values of q. The estimates are $\hat{\lambda}_{0q}$ for the base year and $\hat{\lambda}_{1q}$ for the later year.
2. Draw with replacement from the Q sets of coefficient vectors. The individual draws are denoted $\hat{\lambda}_{0b}$ and $\hat{\lambda}_{1b}$, where b=1,…,B. A uniform distribution is used, i.e., each q is equally likely to be drawn.

3. Draw with replacement from z_{0i} and z_{1j}, where z_{0i} is the vector of explanatory variables for observation i in period 0 (i=1,…,n_0) and z_{1j} is the vector for observation j in period 1 (j=1,…,n_1). Each observation is equally likely to appear in the new vectors, z_{0b} and z_{1b}, b=1,…,B. (Note that the new implied matrices for Z_0 and Z_1 are both $B \times k$ rather than the original $n_0 \times k$ and $n_1 \times k$, where k is the total number of explanatory variables.)

4. Calculate $z_{0b}\hat{\lambda}_{0b}$, $z_{1b}\hat{\lambda}_{1b}$, and $z_{0b}\hat{\lambda}_{1b}$.

5. Estimate the density functions for $z_{0b}\hat{\lambda}_{0b}$, $z_{1b}\hat{\lambda}_{1b}$, and $z_{0b}\hat{\lambda}_{1b}$. The estimates are denoted $\hat{f}_{00}$, $\hat{f}_{11}$, and $\hat{f}_{01}$.

The density function estimates are then used to decompose the overall change in the distribution of predicted house prices, $\hat{f}_{11} - \hat{f}_{00}$, as follows:

$$\left(\hat{f}_{11} - \hat{f}_{00}\right) = \left(\hat{f}_{11} - \hat{f}_{01}\right) + \left(\hat{f}_{10} - \hat{f}_{00}\right) \qquad (2)$$

where $\hat{f}_{11} - \hat{f}_{01}$ is the portion due to changes in the distribution of the explanatory variables and $\hat{f}_{10} - \hat{f}_{00}$ is the portion associated with changes in the estimated coefficients.

The decomposition can also be calculated for individual variables or groups of variables. Since $Z\lambda = \alpha + X\beta + D\gamma$, we can easily simulate groups of variables by defining empirical counterparts to the expression in step 4, such as $\hat{\alpha}_{0b}$, $\hat{\alpha}_{1b}$, $x_{0b}\hat{\beta}_{0b}$, $x_{1b}\hat{\beta}_{1b}$, $d_{0b}\hat{\gamma}_{0b}$, and $d_{1b}\hat{\gamma}_{1b}$, where lower case letters denote a single row of the larger matrix of explanatory variables. To analyze the effects of the structural characteristics, for example, we can calculate the density functions for $\hat{\psi}_{1b} + x_{1b}\hat{\beta}_{1b}$, $\hat{\psi}_{1b} + x_{0b}\hat{\beta}_{0b}$, and $\hat{\psi}_{1b} + x_{1b}\hat{\beta}_{0b}$, where $\hat{\psi}_{1b} = \hat{\alpha}_{1b} + d_{1b}\hat{\gamma}_{1b}$. By maintaining the intercepts and the values of $D\gamma$ at their period one values, these expressions isolate the effects of the structural characteristics. If we again denote the implied density function estimates by $\hat{f}_{00}$, $\hat{f}_{11}$, and $\hat{f}_{01}$, equation (2) forms the base for the decomposition of the effects of changes in the distribution of the structural variables and the distribution of their coefficients on the house price distribution. Directly analogous calculations make it possible to isolate the effects of the location variables (*D*) or any single explanatory variable. It is also possible to isolate the effects of changes in the intercepts over time while holding all explanatory variables constant (although in this case the entire effect is obviously due to a change in the coefficients rather than to a change in the distribution of a variable).

In contrast to the standard Oaxaca (1973) decomposition, equation (2) traces out changes in the entire distribution of house prices. An increase in the price of high-priced relative to low-priced homes could be explained by several factors. First, it could simply be that more homes sold in period 1 in expensive neighborhoods, i.e., the distribution was dominated by homes in places with certain values of *D* that happen to have had high values of λ in both periods. Second, it could be that the distribution of sales across neighborhoods did not change, but the premium associated with relatively expensive neighborhoods – the values of γ – increased significantly. Or it is possible that the return to certain structural characteristics such as square footage increased, or that the period 1 sample happened to be dominated by larger homes. Finally, the change could be explained simply by a relatively larger increase in the intercepts at higher quantiles, i.e., for reasons entirely unrelated to the measurable variables or their coefficients, prices increased more over time in the upper end of the price distribution.

3. Data and Empirical Methods

The agency responsible for conducting reviews of assessment practices in Illinois, the Illinois Department of Revenue (IDOR), provided the data on sales prices. The basic IDOR data comprises sales prices and dates, along with the unique parcel identification number. I merged data on sales in Chicago for 1995 and 2005 with data from the Cook County Assessor's Office to obtain the property's address, lot size, and various structural characteristics. Restricting the sample to sales of single-family homes in Chicago with no missing variables results in 9,873 observations for 1995 and 15,760 observations in 2005. The structural characteristics include standard variables such as total interior area; the number of rooms, bedrooms, and bathrooms; the age of the structure; and dummy variables indicating whether the home has central air conditioning, a fireplace, is built of bricks, and has a one or two-car garage. In addition, I geocoded the addresses to assign each home to one of Chicago's 77 community areas. Although the community area boundaries date to the 1930s, they remain Chicago's accepted definition of neighborhood and the names given to them then remain in common use today.[1] I use the Bureau of Labor Statistics' Chicago metropolitan area CPI to express all sales in 2005 prices. Table 1 presents descriptive statistics for both years.

Throughout the paper, I use a simple kernel density function with a constant bandwidth, *h*, to calculate empirical distributions. The estimated density function at a target point *x* is

[1] Some of the neighborhoods (such as the predominantly commercial Loop district and the area around O'Hare airport) include very few single-family homes. I restricted the sample to community areas having at least 10 sales in both 1995 and 2005. This restriction reduces the number of community areas in the final sample to 67.

$$\hat{f}_0(x) = \frac{1}{n_0 h_0} \sum_{i=1}^{n_0} K\left(\frac{x_{10} - x}{h_0}\right) \tag{3}$$

in period 0 and

$$\hat{f}_1(x) = \frac{1}{n_1 h_1} \sum_{i=1}^{n_1} K\left(\frac{x_{1i} - x}{h_1}\right) \tag{4}$$

in period 1. The change in densities is simply the difference:

$$\Delta(x) = \hat{f}_1(x) - \hat{f}_0(x) \tag{5}$$

I calculate the densities at 400 target values of x ranging from $\min(x_0, x_1)$ to $\max(x_0, x_1)$. This approach produces a smooth estimate of the density functions and the changes in the densities over time.[2]

The density functions for the natural log of sales prices in 1995 and 2005 are shown in Figure 1. The real sales price distribution shifted far to the right in 2005. The distribution was sharply skewed in 1995, with far more low-priced sales than sales from the right side of the distribution. The right tail is clearly thicker in 2005. Figure 2 shows the implied cumulative density function. The horizontal distance between the two cumulative densities is larger at higher percentiles. For example, the 10th percentile of the distribution occurs at a (log) price of 11.235 in 1995, compared with 11.563 in 2005 – a difference of 0.368. The 90th percentile of the distribution occurs at 12.320 in 1995, compared with 12.953 in 2005, or a difference of 0.633. This pattern implies that the rate

---

[2] Since experimentation with alternative bandwidths and kernels had little effect on the overall results, I simply used the defaults in Stata: $h_j = .9 m_j n_j^{-.2}$, where j=0,1; $m_j = \min\left(\sqrt{\text{var}(x_j)},\ \ r_j/1.349\right)$; and r_j is the interquartile range of x_j. This default is a slight variant of the rule of thumb suggested by Silverman (1986). I use the default kernel, $K(u) = \frac{3}{4}\left(1 - .2u^2\right)/\sqrt{5}$ for $|u| < 0$ and $K(u) = 0$ otherwise.

of price appreciation was higher at higher percentiles. Though real sales prices increased throughout the price distribution between 1995 and 2005, the rate of appreciation was greatest for high-priced homes.

4. Regression Results

The base OLS regression results are shown in the first two columns of Table 2. The results are standard: in both 1995 and 2005, house prices increase with lot size, building area, and the number of bathrooms; and prices decline with age. Prices are higher for brick homes with air conditioning, a fireplace, and a garage. After controlling for building area, the results are somewhat mixed for the number of rooms and bedrooms: dividing a given area into smaller rooms does not necessarily increase a home's sales price. Including the controls for community area fixed effects, the regressions explain a respectable 69% of the variation in the log of house price in 1995 and 75% in 2005.

In results not reported here, the *t*-statistics for interaction terms between the 2005 dummy variable and the other explanatory variables imply that the differences between the 1995 and 2005 coefficients are statistically significant at the 5% level for building area, the number of rooms, central air conditioning, brick construction, both garage variables, and age. The differences are not statistically significant for building area, the number of bedrooms, the number of bathrooms, and the presence of a fireplace. In terms of economic significance, it appears that the price of an additional square foot of building area and the value of a garage fell between 1995 and 2005, while the value of central air conditioning increased. The discount associated with an additional ten years of age declined from -2.19% in 1995 to -1.12% in 2005. The importance of neighborhood fixed effects also appears to have increased over time: the *F*-statistic for identical values for all 67 neighborhood fixed effects increases from 203.207 in 1995 to 511.342 in 2005.

To calculate the Machado-Mata (2005) decomposition, I estimated 49 quantile regressions for quantiles ranging from q=.02 to .98 in increments of .02. Table 2 shows the estimates for representative (25% and 50% and 75%) quantiles.[3] Scatter plots of the regression coefficients by quantile are shown in Figure 3 for the eleven structural variables. Several variables exhibit significant quantile effects. For 1995 sales, an additional unit of building area adds much more to price at high quantiles. In 2005 the quantile effects for building area have largely disappeared, with perhaps a slight reversal of the former pattern. In 1995, dividing a house into more rooms tended to produce lower prices at lower quantiles, while in 2005 this variable did not have a statistically significant effect on prices. The number of bedrooms did not have a statistically significant effect in the quantile regressions in 2005, while a larger number of bedrooms increased prices at low quantiles in 1995. More bathrooms increased house prices by more at higher price quantiles in both years. Quantile effects are evident for central air conditioning in 2005, adding more to house prices at lower quantiles. Garages added more to house prices at lower quantiles in both years, but the implicit prices fall in 2005. In 1995, age tends to decrease prices at all quantiles. In 2005, the deleterious effects of age are confined to lower price quantiles.

[3] The quantile estimator was proposed by Koenker and Bassett (1978), and has recently increased in popularity following the work of Buchinsky (1994, 1998a, 1998b, 2001). Buchinsky (1998b), Koenker (2005), and Koenker and Hallock (2001) present useful surveys. Examples for real estate pricing include Gyourko and Tracy (1999), and McMillen and Thorsnes (2006), and Thorsen (1994).

5. The Decompositions

The distributions of the log of sales price in 1995 and 2005 were shown in Figure 1. The solid line in Figure 4 shows the change in the distribution over the decade. The number of sales with high real sales prices rises while the number of sales with low prices falls. How much of this change was due to a change in the distribution of the explanatory variables – i.e., the sample composition – and how much was due to a change in the coefficients of the hedonic price functions? To address this question, I make 5000 independent draws from the rows of the 1995 and 2005 explanatory variable matrices and 5000 independent draws from the estimated quantile coefficient vectors for the two years. Following the procedures outlined in Section 2, I use the results to construct the estimates $Z\hat{\lambda}_q$ using various combinations of 1995 and 2005 values of Z and $\hat{\lambda}_q$, and then estimate the implied densities using equations (3) and (4). Equation (5) then provides estimates of the difference in these counterfactual densities.[4]

The dashed lines in Figure 4 show the change in densities when (1) $Z\hat{\lambda}_q$ is evaluated at 2005 levels of $\hat{\lambda}_q$ but Z changes over time (the change due to changes in the explanatory variables), and (2) the estimates are evaluated using 1995 levels of Z while $\hat{\lambda}_q$ changes over time (the change due to changes in the quantile regression coefficients). Figure 4 suggests that virtually the entire change in the distribution of sales prices is due to changes in the coefficients over time. Although homes sell in different neighborhoods at different times, the average home aged 10 years, and the mix of large and small homes

[4] Note that the number of observations for all simulations is the same, so the bandwidth *h* varies only because the variances or interquartile ranges may differ slightly.

changes over time, the primary determinant of the change in the distribution of house prices is simply that the estimated coefficients changed.

If all homes shared equally in an overall increase in real prices, the distribution of real prices would shift to the right even if no other coefficients changed. Thus, this decomposition may be somewhat unfair because it is, in effect, much harder for the explanatory variables to alter the distribution than the coefficients. In addition, it is interesting to know how much of the change in the distribution is due to changes in the coefficients and values of the structural variables and how much is due to changes in where the sales are located. The three panels of Figure 5 repeat the exercise by decomposing the changes into five effects: changes in (1) the intercepts, (2) the distribution of the structural variables, (3) the coefficients for the structural variables, (4) the distribution of sales across neighborhoods, and (5) the coefficients for the neighborhood fixed effects.

Figure 5 suggests that the intercept has more effect on the change in distributions than the explanatory variables. The peak of the change in densities is about 0.50 for the intercepts, compared with less than 0.20 for the structural variables and 0.08 for the neighborhood variables. Moreover, the shape of the function for the intercepts is similar to the original change in the distribution of sales prices, with the intercept shifts leading to a larger number of high-priced sales and fewer low-priced sales. In contrast, the effect of the structural variables is to *reduce* the number of high-priced sales while increasing the number of sales in the middle of the price distribution. We shall see that much of this change is due to changes in the coefficients of building area, which placed a lower premium on larger structures in 2005 compared with 1995. Indeed, Figure 5 shows that most of this structure-induced change in the distribution of sales prices is due to changes in the coefficients. Neighborhood fixed effects have much less influence on the change in the sales price distribution. A tendency for coefficient changes to produce more sales in the left side of the price distribution is largely offset by changes in the distribution of sales across neighborhoods reflecting more sales in high-priced locations.

It might be expected beforehand that a relative increase in the price of high-priced homes would be caused by an increase in the number of homes sold in high-priced areas, an increase in the location premium associated with high-priced area, an increase in the number of homes sold with high-priced structural attributes, or an increase in the return to certain structural attributes. None of these factors appears to account for the change in the house price distribution. Instead, prices appreciated more in high-priced areas because the intercepts of the quantile regressions increased more at high quantiles.

6. Individual Structural Characteristics

Changes in the house price distribution can also be decomposed into the portions due to changes in the distribution of individual variables and their coefficients. The results for the structural variables are shown in Figure 6. Although many of the changes are small and noisy, a few results stand out. First, note that the coefficients for the log of building area and the number of bedrooms changed in a way that would tend to reduce the number of high-priced sales while increasing the number of sales in the middle of the distribution. In contrast, the coefficients for the number of rooms and age changed in a manner similar to the overall change in the house price distribution. All of the decompositions are dominated by changes in the coefficients. Any changes in the distributions of explanatory variables themselves have little or no effect on changes in the house price distribution.

The quantitative significance of these changes can be assessed using the values for $\Delta(x) = \hat{f}_1(x) - \hat{f}_0(x)$. Thus, it is clear from Figure 6 that building areas have more effect on changes in the house price distribution than the number of rooms because $|\Delta(x)|$ reaches values approaching 0.20 for the former while it only approaches 0.04 for the latter. This observation suggests a measure of the relative importance of a variable's contribution to the change in the price distribution: we can compare the average value of $|\Delta(x)|$ across all target values of x to the average value for the total change in the distribution as shown in Figure 4.

The results of these calculations are shown in Table 3. To interpret the results, note that the average value of $|\Delta(x)|$ is 0.222 for the total change in the house price distribution. When this change is decomposed into the portion due to changes in all explanatory variables and their coefficients, the average value of $|\Delta(x)|$ is 6.11% of 0.222 for the explanatory variables, compared with 102.01% for the coefficients.[5] This result re-emphasizes the point that nearly all of the change in the house price distribution is caused by changes in the coefficients. The most important source of change is clearly the intercept shifts: changes in the estimated coefficients produce an average value of $|\Delta(x)|$ that is 128.31% higher than the value for the overall distribution. The relative values of $|\Delta(x)|$ are high (defined arbitrarily as near 10% or higher) only for the

[5] The percentage can exceed 100% when the changes in the distribution associated with the explanatory variables and their coefficients move in opposite directions.

coefficients for the full set of structural variables, the full set of location variables, lot size, building area, and age. By this measure, changes in the distribution of the explanatory variables themselves never account for a substantial portion of the change in the house price distribution.

7. Conclusion

This study differs from most previous work on housing by focusing on determinants of changes in the full distribution of prices. The house price distribution for sales of single-family homes in Chicago was highly skewed in 1995, with only a small number of sales of very high-priced homes. Though the entire distribution of *real* prices shifted to the right in 2005, the increase was larger for high-priced homes. I use a procedure developed by Machado and Mata (2005) to decompose this change in the house price distribution into the portion due to changes in the distributions of the explanatory variables and their coefficients. Although it might be expected beforehand that the change in the distribution would be explained by an increase in the number of sales in high-priced neighborhoods or the number of homes sold with comparatively expensive structural characteristics, it turns out that nearly all of the change is explained by changes in the estimated coefficients. Most importantly, the increase in the intercepts of quantile hedonic house price regressions is higher at higher quantiles. Location and housing attributes do not explain the change in the house price distributions; rather, there has been a shift upward in the hedonic price functions that lead to higher-priced homes being even more highly priced than before.

References

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.

Dickey, Heather, "Regional Earnings Inequality in Great Britain: Evidence from Quantile Regressions," *Journal of Regional Science* 47 (2007), 775-806.

Gosling, Amanda, Stephen Machin, and Costas Meghir, "The Changing Distribution of Male Wages in the U..K," *Review of Economic Studies* 67 (2000), 635-666.

Gyourko, J. and J. Tracy, "A Look at Real Housing Prices and Incomes: Some Implications for Housing Affordability and Quality," *Federal Reserve Bank of* New York Policy Review (1999), 63-77.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101 (1993), 410-442.

Katz, Lawrence F. and Kevin M. Murphy, "Changes in the Structure of Wages 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics* 107 (1992), 35-78.

Koenker, Roger, *Quantile Regression*, Cambridge University Press, New York (2005).

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.

Machado, Jose A. F. and Jose Mata, "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression," *Journal of Applied Econometrics* 20 (2005), 445-465.

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. and Paul Thorsnes, “Housing Renovations and the Quantile Repeat Sales Price Index, *Real Estate Economics* 34 (2006), 567-587.

Oaxaca, Ronald L., “Male-Female Differentials in Urban Labor Markets,” *International Economic Review* 14 (1973), 693-709.

Silverman, A. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York (1986).

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
Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
1995 (9,873 observations)				
Log of sales price	11.807	0.465	9.978	13.220
Log of lot size	8.274	0.314	6.252	10.731
Log of building area	7.087	0.301	5.991	8.550
Number of rooms	5.541	1.357	2	12
Number of bedrooms	2.890	0.775	1	7
Number of bathrooms	1.338	0.521	1	5.5
Central air conditioning	0.224	0.417	0	1
Fireplace	0.097	0.296	0	1
Brick construction	0.646	0.478	0	1
Garage, 1-car	0.287	0.452	0	1
Garage, 2^{+}-car	0.492	0.500	0	1
Age	62.242	24.610	0	139
2005 (15,620 observations)				
Log of sales price	12.288	0.558	10.7	13.9
Log of lot size	8.237	0.319	6.3	10.1
Log of building area	7.048	0.306	6.0	8.9
Number of rooms	5.449	1.398	2	12
Number of bedrooms	2.837	0.792	1	8
Number of bathrooms	1.316	0.507	1	5.5
Central air conditioning	0.185	0.389	0	1
Fireplace	0.087	0.282	0	1
Brick construction	0.586	0.493	0	1
Garage, 1-car	0.297	0.457	0	1
Garage, 2^{+}-car	0.453	0.498	0	1
Age	75.810	27.150	5	189

Table 2
Regression Results

	OLS		25%		50%		75%	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
1995								
Log of lot size	0.243	0.011	0.239	0.013	0.247	0.009	0.264	0.010
Log of building area	0.303	0.014	0.262	0.016	0.288	0.012	0.296	0.014
Number of rooms	-0.008	0.003	-0.013	0.004	-0.005	0.003	-0.002	0.003
Number of bedrooms	0.012	0.006	0.027	0.006	0.011	0.005	-0.001	0.005
Number of bathrooms	0.037	0.007	0.031	0.008	0.049	0.006	0.068	0.007
Central air conditioning	0.005	0.007	0.003	0.008	-0.008	0.006	-0.002	0.007
Fireplace	0.054	0.010	0.037	0.012	0.061	0.008	0.073	0.010
Brick construction	0.069	0.007	0.060	0.007	0.036	0.005	0.017	0.006
Garage, 1-car	0.066	0.008	0.073	0.008	0.050	0.006	0.034	0.007
Garage, 2^+-car	0.074	0.007	0.087	0.008	0.069	0.006	0.048	0.007
Age	-0.002	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
Constant	7.636	0.114	7.323	0.153	7.134	0.110	6.955	0.132
R^2	0.690							
2005								
Log of lot size	0.252	0.009	0.252	0.011	0.246	0.007	0.242	0.008
Log of building area	0.248	0.012	0.260	0.013	0.245	0.009	0.245	0.010
Number of rooms	0.004	0.003	-0.003	0.003	0.002	0.002	0.002	0.003
Number of bedrooms	0.001	0.005	0.006	0.006	0.001	0.004	0.002	0.004
Number of bathrooms	0.041	0.006	0.039	0.007	0.043	0.005	0.049	0.005
Central air conditioning	0.029	0.007	0.017	0.008	0.014	0.006	0.007	0.006
Fireplace	0.069	0.009	0.053	0.011	0.071	0.008	0.073	0.008
Brick construction	0.043	0.005	0.040	0.006	0.033	0.004	0.029	0.005
Garage, 1-car	0.034	0.006	0.048	0.007	0.031	0.005	0.020	0.006
Garage, 2^+-car	0.042	0.006	0.063	0.007	0.043	0.005	0.028	0.005
Age	-0.001	0.000	-0.002	0.000	-0.001	0.000	-0.001	0.000
Constant	8.403	0.095	7.844	0.119	8.191	0.082	8.527	0.088
R^2	0.752							

Note. The dependent variable is the natural log of sales price. The regressions also include 66 community area dummy variables.

Table 3
Decomposition of Distributional Changes

	Variables	Coefficients
All Explanatory Variables	0.0611	1.0201
Intercept	0.0000	1.2831
Structural Variables	0.0611	0.2895
Location Variables	0.0650	0.1708
Log of lot size	0.0265	0.0996
Log of building area	0.0201	0.3413
Number of rooms	0.0023	0.0745
Number of bedrooms	0.0014	0.0545
Number of bathrooms	0.0078	0.0351
Central air conditioning	0.0024	0.0051
Fireplace	0.0054	0.0033
Brick construction	0.0054	0.0149
Garage, 1-car	0.0043	0.0128
Garage, 2^{+}-car	0.0099	0.0267
Age	0.0271	0.1459

Figure 1
Kernel Density Estimates for Log of Real Sales Price

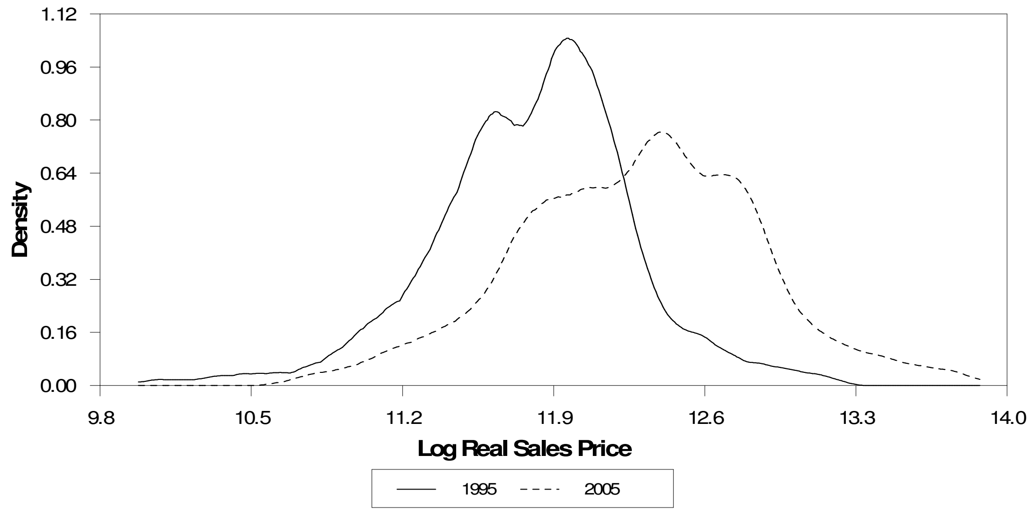


Figure 2
Estimated Cumulative Density Function for Log of Real Sales Price

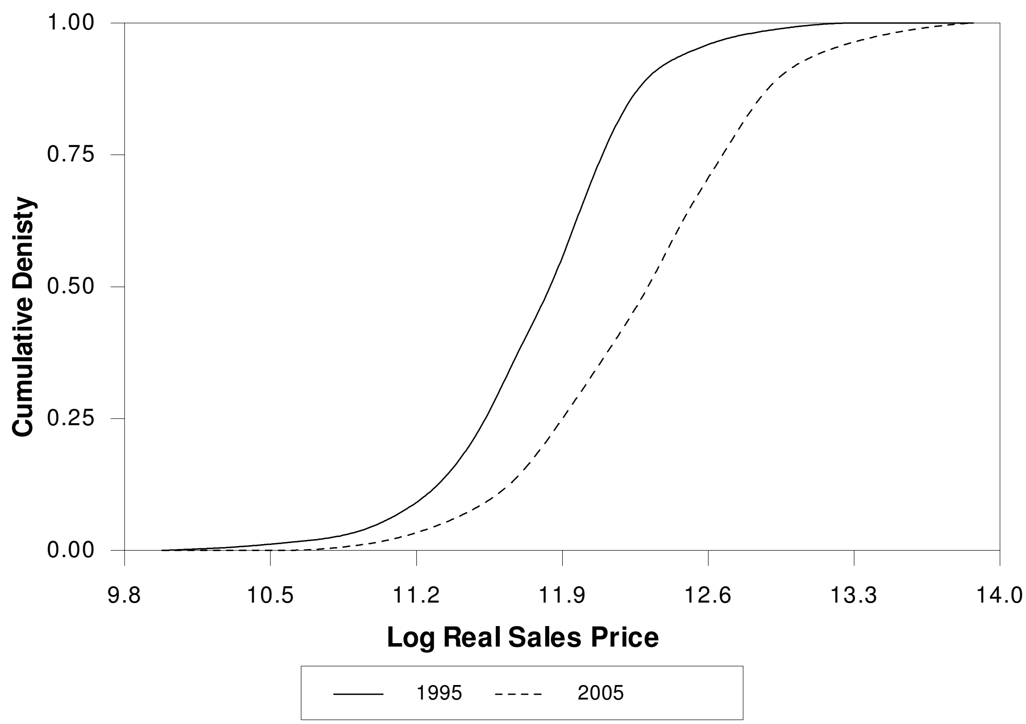
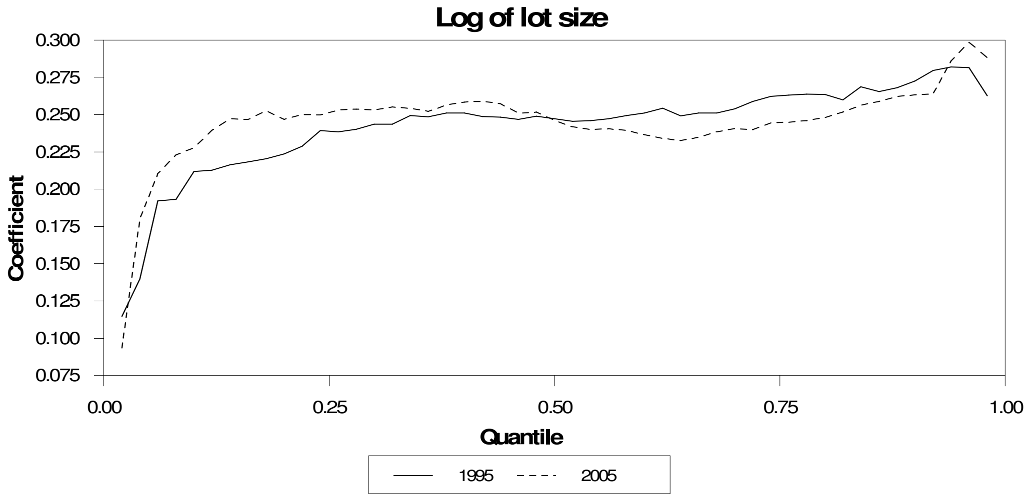
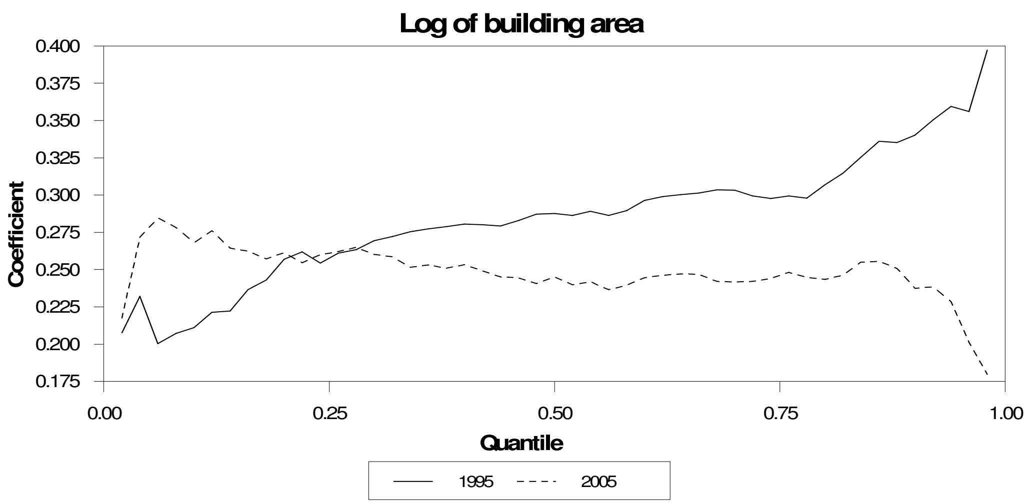
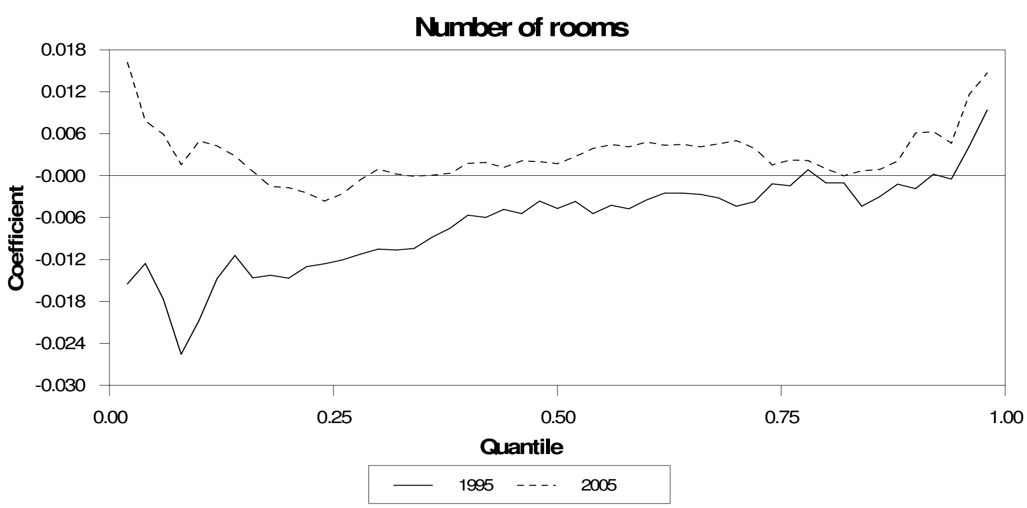
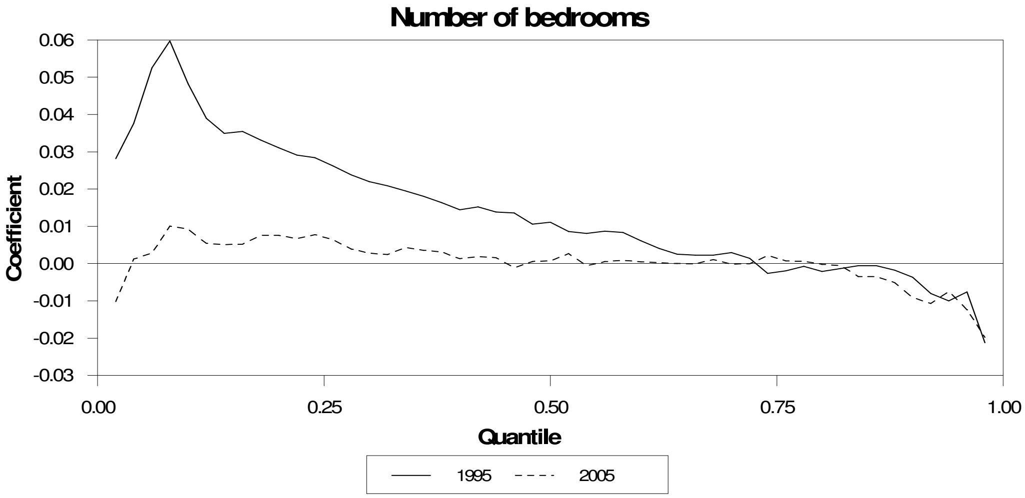


Figure 3
Coefficient Estimates by Quantile

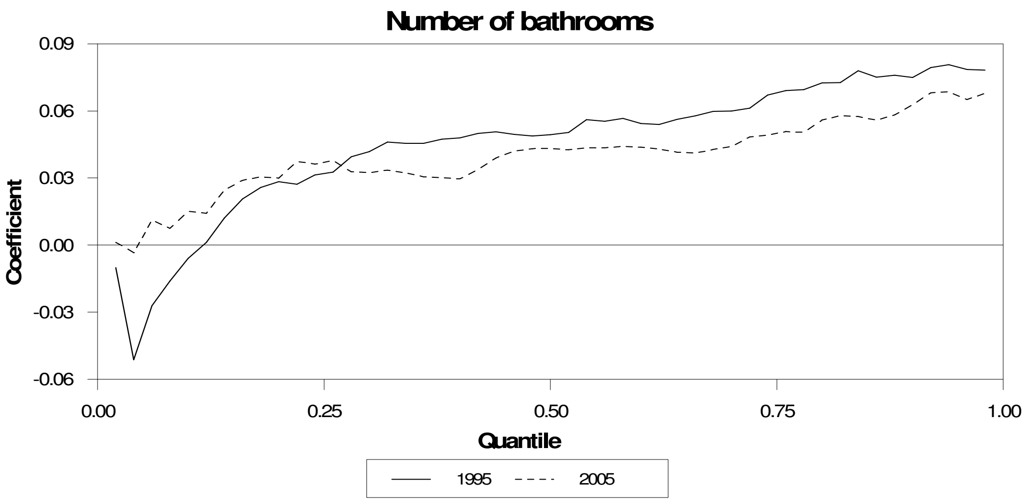




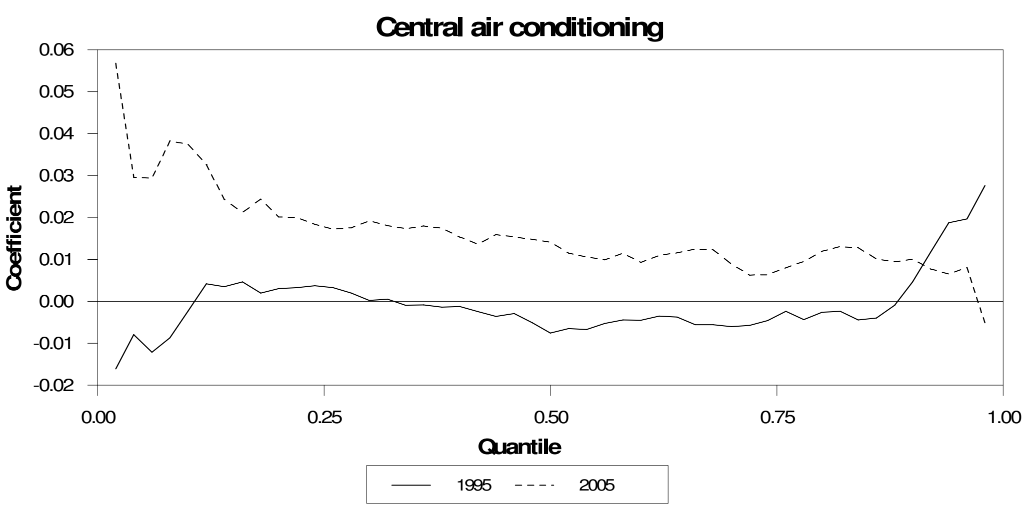




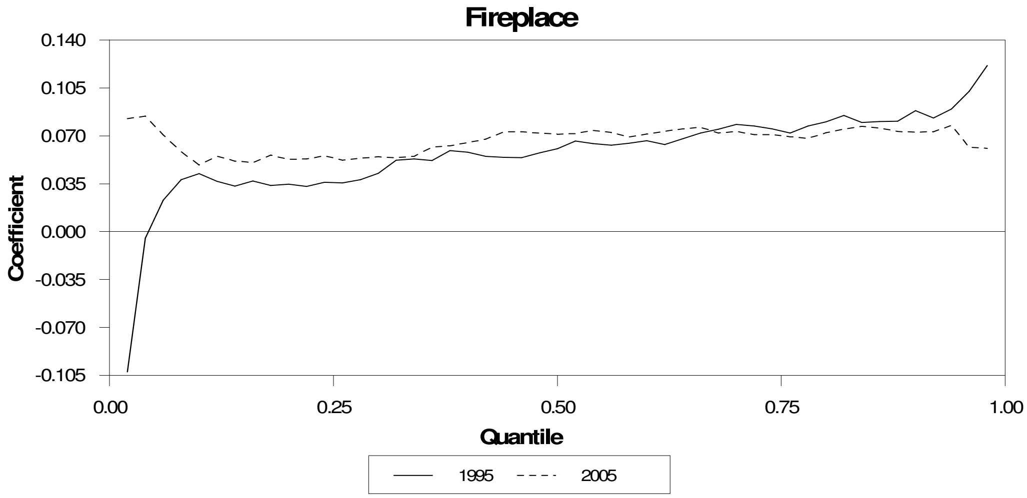
Number of bedrooms
Coefficient
Quantile
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0.00
0.25
0.50
0.75
1.00
1995
2005

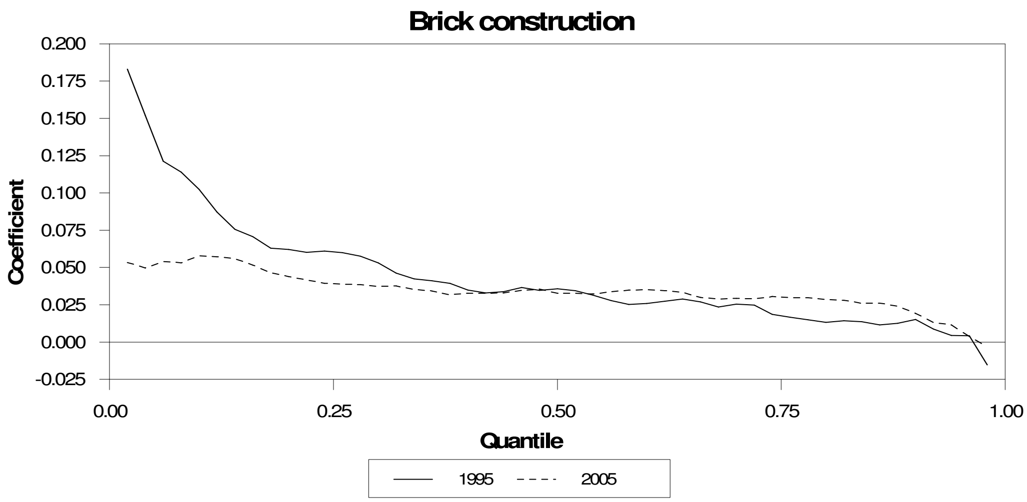


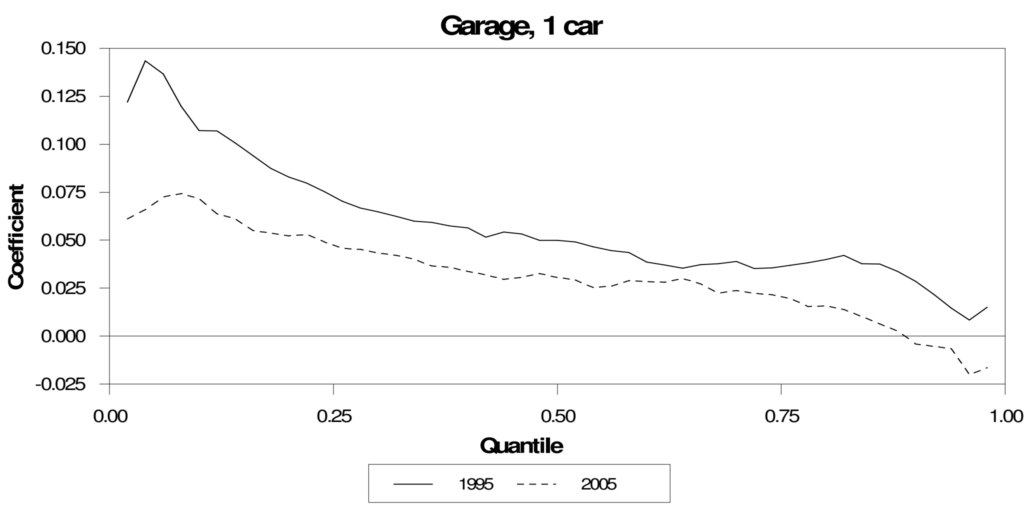
Number of bathrooms
Coefficient
Quantile
0.09
0.06
0.03
0.00
-0.03
-0.06
0.00
0.25
0.50
0.75
1.00
1995
2005

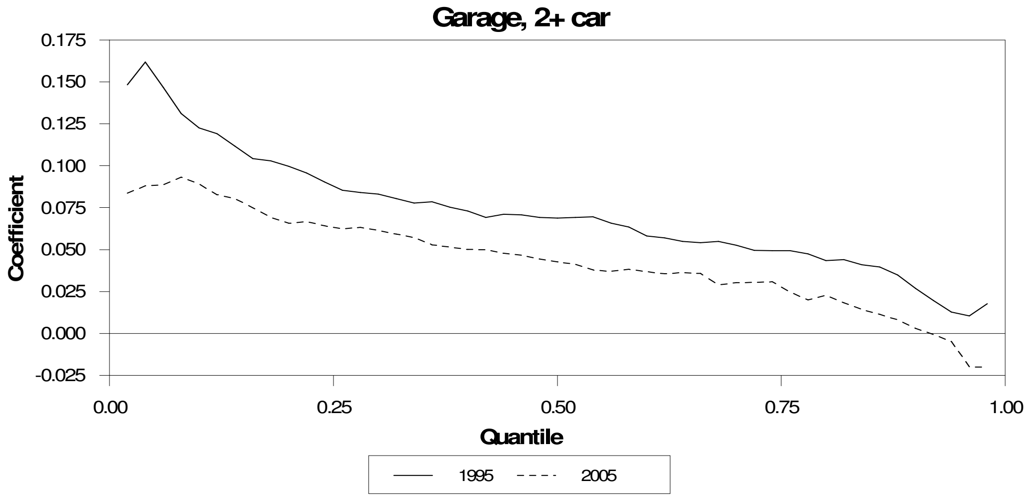


Central air conditioning
Coefficient
Quantile
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
0.00
0.25
0.50
0.75
1.00
1995
2005

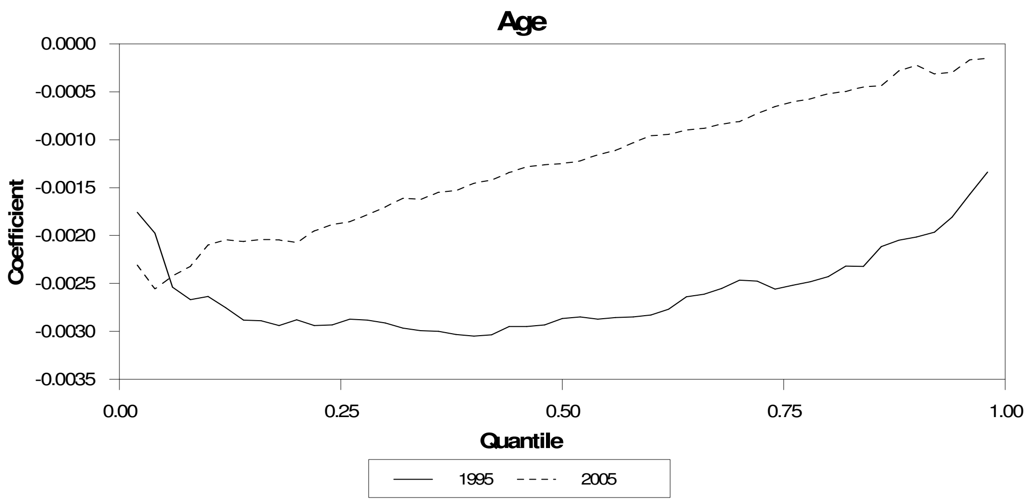

Fireplace
0.140
0.105
0.070
0.035
0.000
-0.035
-0.070
-0.105
Coefficient
0.00
0.25
0.50
0.75
1.00
Quantile
1995
2005


Brick construction
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
Coefficient
0.00
0.25
0.50
0.75
1.00
Quantile
1995
2005


Garage, 1 car
0.150
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
Coefficient
0.00
0.25
0.50
0.75
1.00
Quantile
1995
2005



Garage, 2+ car
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
Coefficient
0.00
0.25
0.50
0.75
1.00
Quantile
1995
2005



Age
0.0000
-0.0005
-0.0010
-0.0015
-0.0020
-0.0025
-0.0030
-0.0035
Coefficient
0.00
0.25
0.50
0.75
1.00
Quantile
1995
2005

Figure 4
Decomposition of Density Changes

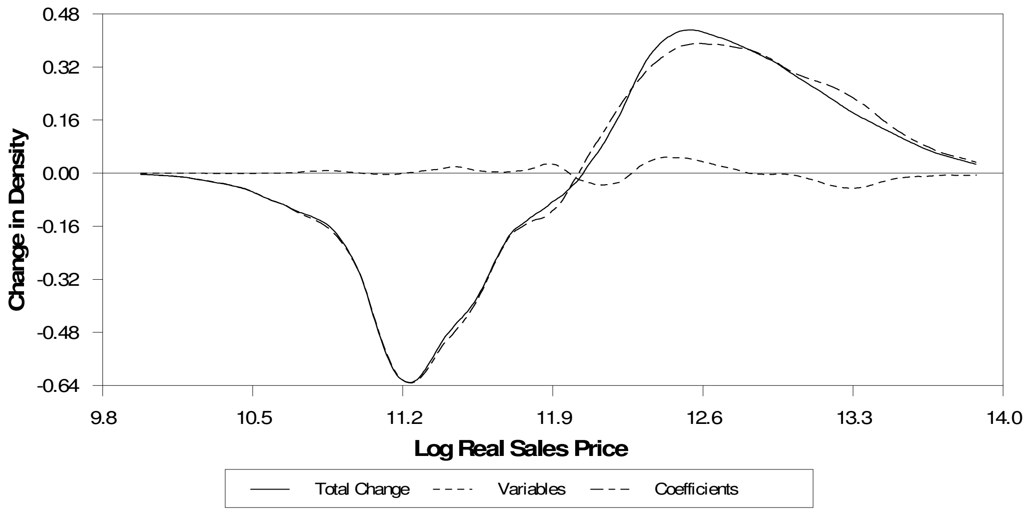


Figure 5
Decomposition of Density Changes by Groups of Variables

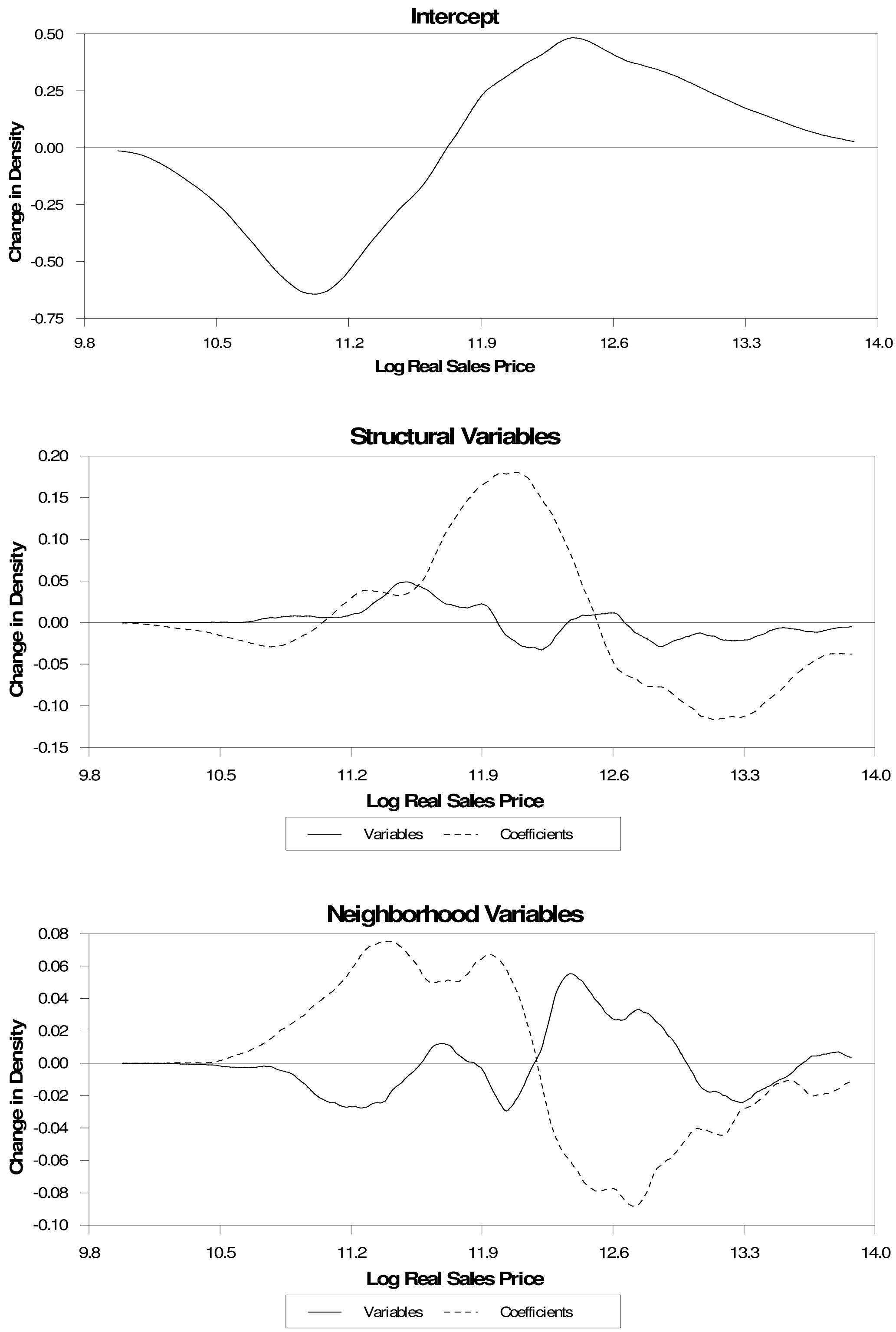
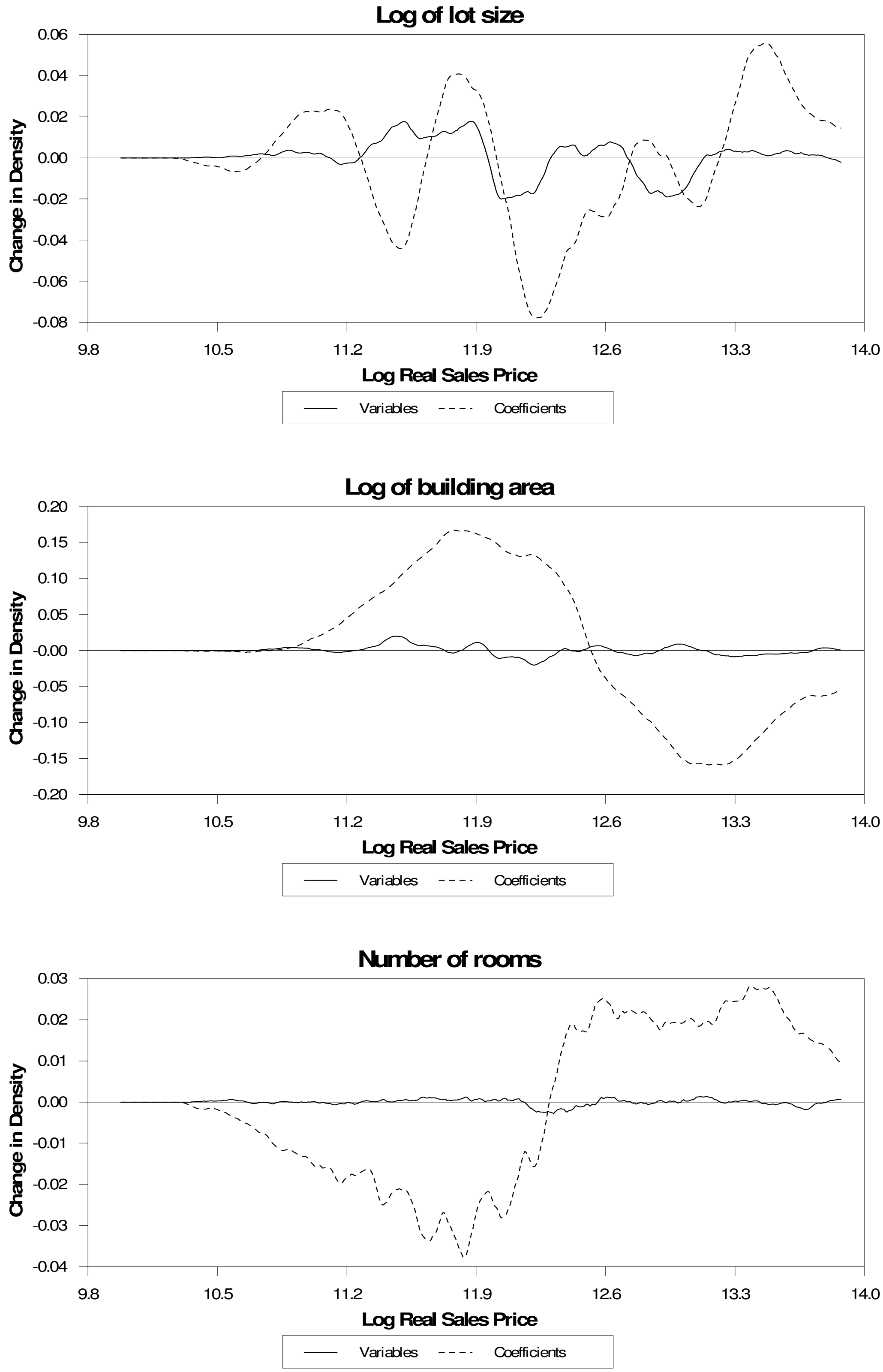
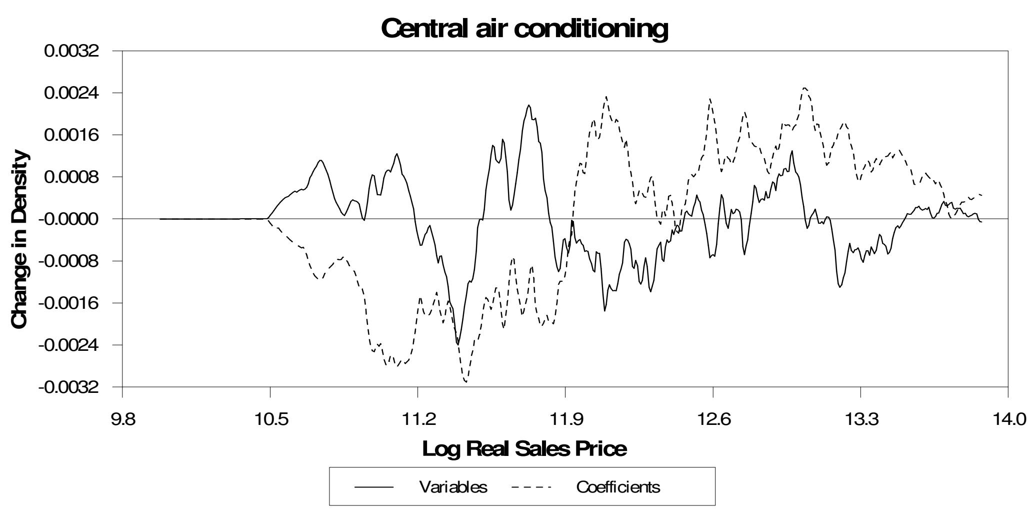


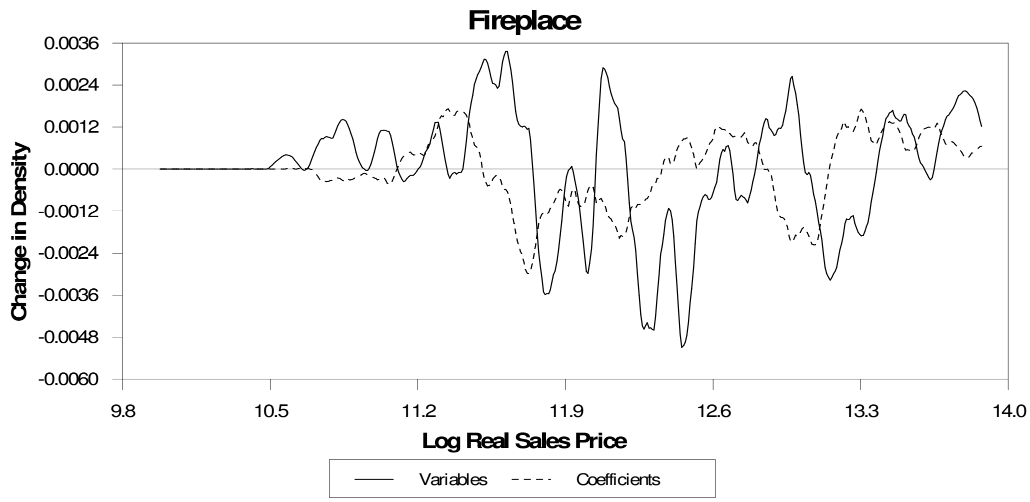
Figure 6
Decomposition of Density Changes: Individual Structural Variables

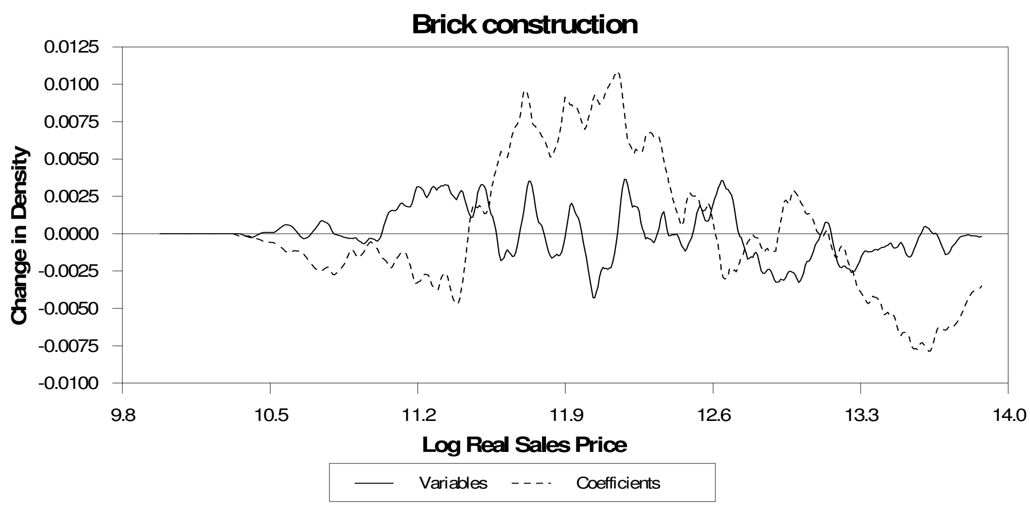


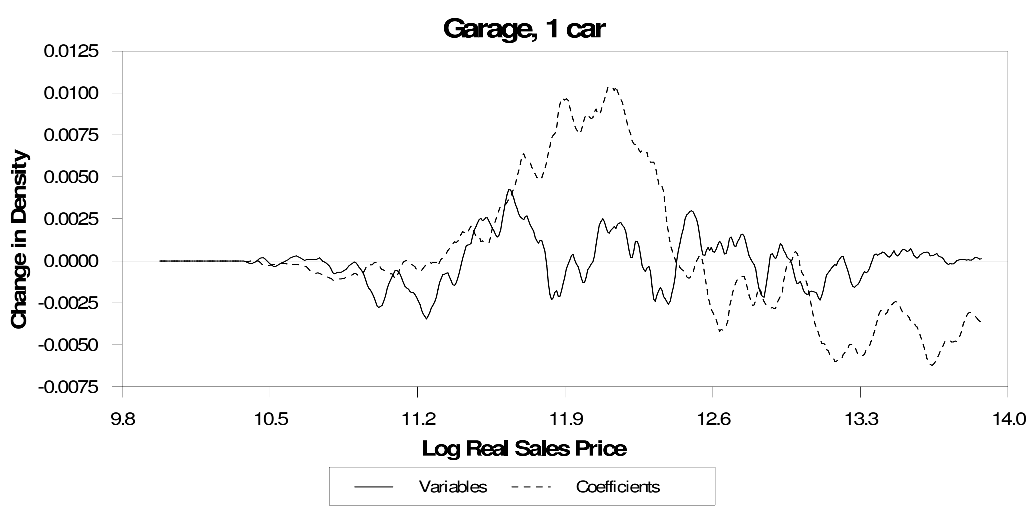

Number of bedrooms
Change in Density
0.032
0.024
0.016
0.008
0.000
-0.008
-0.016
-0.024
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients

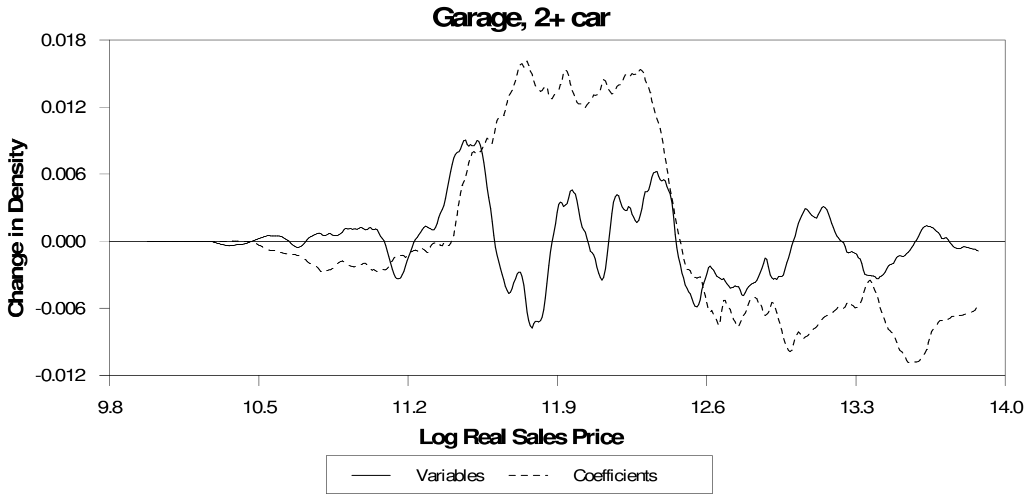

Number of bathrooms
Change in Density
0.021
0.014
0.007
-0.000
-0.007
-0.014
-0.021
-0.028
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Central air conditioning
Change in Density
0.0032
0.0024
0.0016
0.0008
-0.0000
-0.0008
-0.0016
-0.0024
-0.0032
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Fireplace
0.0036
0.0024
0.0012
0.0000
-0.0012
-0.0024
-0.0036
-0.0048
-0.0060
Change in Density
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Brick construction
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
-0.0025
-0.0050
-0.0075
-0.0100
Change in Density
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Garage, 1 car
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
-0.0025
-0.0050
-0.0075
Change in Density
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Garage, 2+ car
Change in Density
0.018
0.012
0.006
0.000
-0.006
-0.012
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients


Age
Change in Density
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
9.8
10.5
11.2
11.9
12.6
13.3
14.0
Log Real Sales Price
Variables
Coefficients