Local Quantile House Price Indices

Daniel McMillen

Department of Economics
214 David Kinley Hall
1407 W. Gregory
Urbana, Illinois 61801
217-333-8930
mcmillen@illinois.edu

December 9, 2013

Abstract

Locally weighted quantile regressions allow the coefficients of hedonic house price functions to vary over space. Using data on all house sales in Cook County, Illinois, for 2000-2011, I show how the full distribution of appreciation rates changed over time in small geographic areas. The estimates reveal significant spatial variation in appreciation rates both geographically and across the distribution of house prices. During the boom, house prices rose most rapidly among lower-priced homes, particularly on the South and West sides of Chicago. Prices then declined most rapidly afterward in these same areas. In contrast, high-priced homes in the Near North Side of the city and in the far North suburbs had only moderate declines in prices after 2006. The results clearly indicate that standard approaches to estimating house price indices over-simplify what is actually a rich set of spatial and temporal variation in appreciation rates.

1. Introduction

House prices do not necessarily change at a uniform rate throughout an urban area. Prices may decline more slowly in highly desirable, high-priced areas during a recession, and they may appreciate more rapidly in low-priced areas in boom periods. Or the opposite patterns may describe the change in the house price distribution. In either case, a single price index does not adequately describe the changes in prices across the full urban area. Nonetheless, the most commonly used price indices – those based on median prices, repeat sales, and hedonic price functions – all typically are based on an assumption that appreciation rates do not vary across neighborhoods or house prices.

Several authors estimate separate price indices for locations within an urban area. Examples include Archer, Gatzlaff, and Ling (1996); Case and Mayer (1996); McMillen (2003); Meese and Wallace (1991); Monkkonen, Wong, and Begley (2012); Ries and Somerville (2010); Rouwendal and Longhi (2008); Schmitz, Shultz, and Sindt (2008); and Weber, Bhatta, and Merriman (2007). A related literature explicitly attempts to identify housing submarkets within cities (e.g., Bourassa, et al., 1999; Bourassa, Hoesli, and Peng, 2003; and Goodman and Thibodeau, 1998, 2007). All of these papers use regression procedures to estimate a single expected appreciation rate for each submarket under consideration.

Recently, several authors have used quantile estimation procedures to analyze variation in appreciation rates within each sample area. Examples of studies that focus on the distribution of price changes include Cobb-Clark and Sinning (2011); Coulson and McMillen (2007), Deng, McMillen, and Sing (2012, forthcoming); McMillen (2008, 2012a); and Nicodemo and Raya (2012). Other authors have used quantile approaches to estimate hedonic price functions using cross sectional data (Kostov, 2009; Liao and Wang, 2012; Zahirovic-Herbert and Chatterjee, 2012;

and Zeitz, Zeitz, and Sirmans, 2008). Related procedures have also been used to analyze the distribution of times on the market (Carrillo and Pope, 2012).

My objective in this paper is to merge these two traditions to show how the full distribution of prices changed over time in small geographic areas. Following McMillen (2012a), I first use a matching estimator to assure that the overall distribution of housing characteristics does not vary significantly over time. Next, I estimate hedonic price functions using data for all house sales in Cook County, Illinois for 2000-2011. I allow appreciation rates to vary across Chicago and its suburbs by using a locally weighted estimation procedure that places more weights on sales that are closer to a set of target points. This nonparametric estimation procedure allows appreciation rates to vary smoothly over space. I estimate locally weighted quantile regressions for a variety of quantiles. Although the estimation procedure is computer intensive, the results are easy to summarize using kernel density estimates.

The results provide fascinating insights into the rise and fall of house prices over the boom and bust periods of the past decade. While any price index would show that prices rose dramatically prior to the end of 2007 and then fell significantly afterward, the locally weighted quantile estimates reveal significant spatial variation in the estimates, as well as variation across the distribution of house prices. During the boom, house prices rose most rapidly among lower-priced homes, particularly on the South and West sides of Chicago. Prices then declined most rapidly afterward in these same areas. In contrast, high-priced homes along the Lakefront had only moderate declines in prices after 2007. The results clearly indicate that standard approaches to estimating house price indices over-simplify what is actually a rich set of spatial and temporal variations in appreciation rates.

2. Mean-Based Price Indices

A typical specification of a hedonic price function expresses the natural log of sale price as a function of characteristics of the structure, location, and the sale date. Let *P* represent sale price, and let *X* represent the combination of variables representing characteristics of the structure and the location. Also, let D_t represent a dummy variable indicating a sale at time *t*. The hedonic price function is:

$$lnP_{it} = X_{it}\beta_t + \sum_{t=1}^{T} D_{it}\delta_t + u_{it} \tag{1}$$

An alternative specification imposes the assumption that the coefficients on *X* do not change over time:

$$lnP_{it} = X_{it}\beta + \sum_{t=1}^{T} D_{it}\delta_t + u_{it} \tag{2}$$

The more general specification, Equation (1), is equivalent to estimating a separate equation for each date. In this case, a price index can be constructed by assuming a set of values for *X*. In Equation (2), the price index is simply the set of coefficients on *D*, where *D* is the matrix comprising the full set of *T* dummy variables.

Invoking an additional assumption that X_{it} is constant over time for each property leads to the repeat sale estimator of Bailey, Muth, and Nourse (1963) and Case and Shiller (1989):

$$lnP_{it} - lnP_{is} = \delta_t - \delta_s + u_{it} - u_{is} \tag{3}$$

where s < t. The repeat sales estimator is often presented as a potential solution to the bias that occurs in hedonic estimation when omitted characteristics of the structure or location are correlated with the error term: if the omitted variables and their coefficients are constant over time, they disappear in the transformation from Equation (2) to Equation (3). However, as argued in McMillen (2012a), the real power of the repeat sales estimator lies its restriction of the sample to the set of properties that sold at least twice during the sample period. In an important but overlooked paper, Wang and Zorn (1997) show that repeat sales estimates are identical to period-by-period sample averages for lnP_t when the number of observations is the same for each period within the repeat sales sample. This result holds whether the variables in X_{it} are constant over time or not, and also when some variables are omitted from X_{it}. Once the sample has been restricted to repeat sales, the only thing that is accomplished by estimating Equation (3) is to reweight the period-by-period sample averages according to the number of sales in each period.

McMillen (2012a) notes that Wang and Zorn's (1997) result implies that the repeat sales model is an extreme form of a matching estimator in which a treatment observation – those properties selling at time *t* rather than at time *s* – is matched only with the sale of the same property at another date. Restricting the sample to repeat sales discards what typically is a much larger number of properties that sold only once during the sample period. In addition to a potential loss in efficiency, the result of this restriction is that the repeat sales approach cannot be applied to relatively small geographic areas where the number of repeat sales pairs is likely to be small. McMillen (2012a) suggests using standard matching estimator approaches to pair sales in a base period with similar properties selling at other times. This approach produces significantly larger sample sizes, while still discarding "unusual" sales – those that are not similar to other properties. Averaging over large numbers of similar sales produces accurate measures of the central tendency of sales prices at each time, and the approach is directly comparable to the repeat sales estimator when the number of sales in the matched sample is the same in each period. Alternatively, a hedonic approach can be used to estimate the predicted sale price in each period for a representative home (Equation 2) or to estimate a "quality-controlled" price index (Equation 3).

3. Quantile Price Indices

Standard hedonic and repeat sales price indices focus on the mean sale price. McMillen (2012a) suggests that a quantile approach can be used to estimate an index for any point in the sales price distribution, such as the median, the 10th percentile, or the 90th percentile. The quantile approach can be implemented in several ways. First, just as the averages over time of the predicted values from OLS estimates of Equation (1) return the period-by-period sample averages of lnP_{it}, the average of the predicted values for quantile regression estimates of Equation (1) for quantile q returns the qth percentile of lnP_{it} for each time period. Quantile regression estimates of Equation (1) can also be used to form an index for any quantile q of the sale price distribution using the predicted values for a representative property. Either of these two approaches can also be applied to a matched sample, whether the matches are constructed using repeat sales or through a more general matching approach. Alternatively, Equation (2) can be estimated using a quantile regression if one is willing to impose the assumption that the coefficients on X_{it} are constant for any quantile q (but not necessarily across quantiles).

The quantile and OLS approaches to hedonic estimation of house price indices differ only in that they focus on different points in the house price distribution. Counterparts to the repeat sales price index can be constructed by calculating period-by-period percentiles of the price distribution for repeat sales or other matched samples. Both quantile and OLS approaches can also be applied directly to the repeat sales or matched samples. Just as a standard index shows how the average quality-controlled sale price changes over time, a quantile index shows how the *q*th percentile price evolves over time.

Quantile regressions can also be used to analyze the effects of a change in any explanatory variable, including date of sale, on the overall distribution of sales prices. The approach is discussed in detail in McMillen (2012b). Following the approach that will be used in the empirical section of the paper, suppose we estimate a quantile version of Equation (2) in which the first time indicator variable is dropped and a constant term is included in *X*:

$$Q_{lnP}(q|X_{it}, D_{it}) = X_{it}\beta(q) + \sum_{t=2}^{T} D_{it}\delta_t(q) \tag{4}$$

This notation, which is a slightly modified version of the notation in Koenker (2005), implies that the conditional quantile function for the natural log of sale price at quantile *q* is a linear function of *X* and the set of time of sale indicator variables. The estimated coefficients vary by quantile. A common approach is to estimate Equation (4) at a set of target quantiles, such as *q* = 0.10, 0.25, 0.50, 0.75, and 0.90.

The full distribution of the dependent variable can be traced out by estimating Equation (4) at all possible quantiles. One approach, which is used in a slightly different context by Machado and Mata (2005), is to draw randomly from values of *q* ranging from 0 to 1, and then re-estimating the model each time. If there are *n* observations in the data set and *B* quantiles are drawn, then the dimension of the resulting matrix of predictions is $n \times B$. If the quantile estimates are reasonably smooth across quantiles, then similar results will be obtained by restricting the number of quantiles to a relatively small number, such as *q* = 0.03, 0.05, …, 0.95, 0.97, which implies *B* = 48. If we then collect the predictions into a single vector with *nB* entries, an estimated kernel density function for the predictions will look nearly identical to the density function for the original values of lnP.

This approach can be used to show how the distribution of the dependent variable changes when a single explanatory variable is set at an arbitrary set of values. In the empirical section of the paper, sales dates ranging from 2000 to 2011 are represented by a series of 12 dummy variables ranging from D_0 to D_{11}, and D_0 is omitted from the estimated model. The estimated value of lnP at quantile q in 2000 is simply $X_{it}\hat{\beta}(q)$, and the estimate value at the same quantile in year *t* is $X_{it}\hat{\beta}(q) + \hat{\delta}_t$. The other explanatory variables are set to their observed values, X_{it}. The same set of calculations can be then be conducted for other quantiles. Thus, the results imply *nB* predicted values for $Q_{lnP}(q|D_0 = 1)$ and for $Q_{lnP}(q|D_t = 1)$. Kernel density estimates for these sets of *nB* predicted values show how the distribution of lnP changes when the year of sale changes from 2000 to a later year and the values for X_{it} are set at their actual values. A suitable bootstrap procedure can be used to construct confidence intervals for the counterfactual densities (Chernozhukov, Fernandez-Val, and Melly, forthcoming). A series of density function estimates shows how the full distribution of log sale price changes over time.

4. Locally Weighted Indices

So far, my discussion of alternative approaches for estimating price indices has not been explicitly spatial. In keeping with Equation (1), one approach to allowing for spatial variation in appreciation rates is to estimate a separate hedonic price functions for submarkets within an urban area. Examples of this approach include Archer, Gatzlaff, and Ling (1996); Case and Mayer (1996); Meese and Wallace (1991); Monkkonen, Wong, and Begley (2012); Ries and Somerville

(2010); Rouwendal and Longhi (2008); Schmitz, Shultz, and Sindt (2008); and Weber, Bhatta, and Merriman (2007). An alternative that is in keeping with Equation (2) is to interact a set of neighborhood dummy variables with the set of time dummy variables. The same approaches can also be used for quantile estimation, although estimation times may be high for models with a large number of time and neighborhood interaction variables.

In many cases, it is reasonable to assume that variables such as house prices and appreciation rates vary smoothly over space. Neighborhood dummy variables may produce accurate results if the neighborhoods are defined accurately and regression coefficients change discretely at neighborhood boundaries. An alternative is to use a variant of Cleveland and Devlin's (1988) locally weighted regression (LWR) procedure in which the coefficients of the estimating equation are assumed to vary smoothly over space. This variant, which is often referred to as "geographically weighted regression" or GWR, was used in McMillen (2003) to estimate repeat sales indices that allow appreciation rates to vary smoothly over location within a city. Letting z_{1i} and z_{2i} represent the geographic coordinates (e.g., longitude and latitude or distance north and east of a base location) of the property associated with observation i, the GWR version of the models analyzed here is obtained by writing the coefficients in Equations (1) – (4) as functions of the coordinates. For example, $\beta_t(z_{1i}, z_{2i})$ and $\delta_t(z_{1i}, z_{2i})$ replace β_t and δ_t in Equation (1). Cleveland and Devlin's estimation procedure was introduced to the urban and real estate literature by Meese and Wallace (1991), and the simplified GWR version of the model was first used by McMillen (1996).

Separate locally weighted models are estimated for a set of target locations, with more weight applied to observations that are closer to the target locations. Letting d_i represent the distance from the target site to the location associated with observation i, the weight applied to observation *j* when estimating the model for the target location is $K(d_i)$, where K is any standard kernel weight function. For linear regression models, the estimation procedure is simply a set of weighted least squares regressions, one for each target point. In many studies, each observation serves in turn as a separate target point. However, estimation time can be reduced significantly by taking advantage of the smoothness implied by the LWR approach by interpolating from a smaller set of target points to each location represented in the data set. A detailed discussion is presented in McMillen (2012b).

A similar approach can be used for quantile estimation using results from Chaudhuri (1991), Koenker (2005, chapter 7), and Yu and Jones (1998). Following the notation in Koenker (2005), define the piecewise linear function $\rho_q(u) = u\big(q - I(u < 0)\big)$. The standard quantile approach for a simple regression model involves finding the values for $\hat{\beta}(q)$ that minimize $\sum_i \rho_q(y_i - x_i'\beta)$. Obvious counterparts to this equation can be defined using the notation of Equations (1) and (2). The counterpart to GWR is obtained by finding the values of $\hat{\beta}(q, d)$ that minimize the locally weighted objective function $\sum_i K(d_i)\, \rho_q(y_i - x_i'\beta)$. As is the case for standard GWR models, the idea is simply to place more weight on nearby observations when estimating the model at a target point. Again, estimation time is reduced significantly by interpolating to the full set of data points from a set of target points.

After interpolation, the locally weighted estimation procedure produces a set of $B \times k$ coefficients for each observation, where k is the number of explanatory variables and B again indicates the number of quantiles for which the locally weighted quantile regressions are estimated. Despite this complexity, the estimates can be summarized easily using counterfactual kernel density functions. As before, consider the case where we want to compare predicted values for

2000 to those for another year t. All that differs from before is that the estimated coefficients have a subscript for the individual observation: $Q_{lnP}(q|D_{0i}=1)=X_{it}\hat{\beta}_i(q)$ and $Q_{lnP}(q|D_{ti}=1)=X_{it}\hat{\beta}_i(q)+\hat{\delta}_{ti}$. As before, each set of predictions has $n \times B$ values, and the results can be summarized easily using kernel density estimates.

The locally weighted approach has two important advantages for spatial quantile models when compared with including a potentially large number of neighborhood fixed effects. First, it is based on what often is a more reasonable assumption that spatial effects vary smoothly rather than changing discretely at neighborhood boundaries. Second, it is often the case that some locations have few observations, which leads to imprecise estimates and can lead to problems of convergence for the quantile estimator. Although imprecision is also a problem for ordinary regression models with a large number of fixed effects, the problem is greater for quantile models because more coefficients are estimated since the coefficients vary by quantile.

5. Data

The data set includes all sales of single-family houses in Cook County, Illinois for 2000-2011. With approximately 5.2 million people, Cook County is the second most populous county in the U.S., and it has more residents than all but 21 states. Chicago accounts for 2.7 million of these residents. The Illinois Department of Revenue provided data on sales dates and prices. I then merged this data set with data from the Cook County Assessor's Office showing addresses and a standard set of structural characteristics. The structural characteristics include building area, lot size, the number of rooms, the number of bathrooms, and the year when the home was built. The data set also includes variables indicating whether the home is built of brick and whether it has a basement, central air conditioning, a fireplace, an attic, and a garage. After dropping observations with missing data, the full data set includes 409,994 sales.

Cook County is divided into three assessment districts – the City of Chicago, the North Suburbs, and the South Suburbs. These districts are a natural starting point for defining geographic sub-regions, and empirically it does turn out that changes in the distribution of house prices differ across the three areas. The top panel of Table 1 shows the number of sales annually in each region and in the full data set, while the top panel of Table 2 presents descriptive statistics across time for each district.

As discussed in Section 2, the repeat sales estimator is an extreme version of a matching estimator in which each sale is matched with the sale of the same property at another time. With a full set of standard explanatory variables, a "selection on observables" assumption is reasonable, i.e., $u_t \perp (D_t, X_t)$. If this condition holds and Equation (2) is the correct model specification, then linear regressions provide unbiased estimates of the δ_t. Nonetheless, it may still be preferable to drop sales of relatively unusual properties, such as extremely small or very old homes. A goal of a matching estimator is to pair treatment observations with similar observations from a control group. In the case of a house price index, the control group is the base period, while each subsequent time is the treatment: what would be the expected price of a property if it were to sell in time *t* rather than in the base period? As Ho et al (2007) emphasize, using a matching procedure to pre-process the data is likely to make the estimates less model dependent. McMillen (2012a) finds support for this point: whereas estimated repeat sales, hedonic, and simple median price indices are significantly different before matching, they are nearly the same for matched samples.

As in McMillen (2012a), I use a simple propensity score approach to construct the matched samples. Using 2000 as the base each time, I estimate a series of logit models for each subsequent year with the indicator variable $y_t \equiv I(year = t)$ as the dependent variable and the structural characteristics shown in Table 2 as the explanatory variables. I then use the predicted probability of sale in year *t* to match each year *t* observation to its closest counterpart from 2000. When year *t* has more sales than in 2000, the matched sample for year *t* will have no more than the number of observations in 2000, n_0. Additional observations are dropped if they fall outside the support of the estimated probabilities, i.e., if $\hat{p}(y_t = 1) \notin \left[min\left(\hat{p}(y_t = 0)\right), max\left(\hat{p}(y_t = 0)\right)\right]$, where $\hat{p}$ is the estimated probability that the property sold in year *t* rather than in 2000. In years where the number of observations in year *t* is less than n_0, all observations will remain in the matched sample unless this support condition is violated. Observations are dropped from the 2000 sample only when they fail to find a match with a sale in any of the subsequent years. I estimate separate logit models for Chicago, the North suburbs, and the South suburbs, and I also construct the matches separately by region.

The lower panel of Table 1 shows the resulting number of observations in the matched samples for the three regions. The lower panel of Table 2 presents descriptive statistics for each region across all years. Table 3 shows the variation in the means over time for each of the regions. The means of the explanatory variables are quite similar even in the full sample. The relatively low variation in the means over time suggests that the annual data sets are all reasonably balanced.

6. Estimated Hedonic Price Indices

The results of a standard hedonic regression analysis are presented in Table 2. The set of matched samples serves as the data set for the hedonic estimates, as well as for the quantile estimates to be presented later. Equation (2) serves as the basis for the estimated hedonic price functions. The equation has constant coefficients for the structural coefficients over time, with fixed effects for census tracts and the year of sale. The equations fit the data reasonably well, and the results are as expected. Sales prices increase with building area, lot size, the number of bathrooms. Brick construction, a basement, central air conditioning, a fireplace, and a garage all are associated with higher sales prices. Sales prices are also higher for newer homes.

The coefficients on the year of sale represent the price indices. The indices are displayed in Figure 1. Prices peaked in 2006 in each region after a period of rapid appreciation. The implied annual appreciation rate for 2000-2006 was approximately 7% in both the North and South suburban regions, and it was nearly 9% in Chicago. Prices dropped moderately between 2006 and 2007, after which they fell dramatically, particularly in Chicago. The implied annual rate of change in prices was approximately -21% in Chicago and -16% in the South Suburbs, compared with a somewhat more modest rate of -9.4% in the North Suburbs.

7. Percentiles for Matched Samples

A direct comparison of quantile estimation to the price indices shown in Figure 1 is infeasible due to the large number of census tract fixed effects included in the hedonic price function estimates. In this section, I simply show how the 10%, 50%, and 90% percentiles of the sales price distribution vary over time. The raw percentiles are equivalent to the average predictions at these quantiles for quantile regressions that do not include controls for location. In the next section, I control for location by estimating locally weighted versions of quantile regressions.

Figure 2 shows how the percentiles of the sale price distribution change over time. The drop in prices after 2006 is evident at each percentile. However, it is much more pronounced at the lower end of price distribution, particularly in Chicago and the South Suburbs. To facilitate a comparison of the price changes across regions, the percentiles are transformed into indices in Figure 3 by subtracting the value for 2000 from each year's value. The median index in the second panel of Figure 3 is roughly comparable to the hedonic estimates of Figure 1: prices rose fastest in Chicago up to 2006, but then declined more rapidly than in the other regions. The first panel of Figure 3 shows that this decline was even more marked for the 10% percentile. The 10% percentile of the sale price distribution declined much less in the North Suburbs than in the other two regions. It is also noteworthy that the 90th percentile of the price distribution stopped increasing earlier in the South Suburbs than in the other two regions.

A median price index is comparable to a standard repeat sales index when the medians are constructed using repeat sales or matched samples. The differences between the approaches can be illustrated with a simple two-period model based on Equation (2) with a single explanatory variable, x. In this case, the means of log sale price in periods 1 and 2 are $\bar{y}_1 = \bar{x}_1\beta + \delta_1 + \bar{u}_1$ and $\bar{y}_2 = \bar{x}_2\beta + \delta_2 + \bar{u}_2$, where $y_t \equiv lnP_t$. As noted before, the repeat sales estimator is equivalent to a simple difference in means when the number of sales is the same in both periods. Thus, $\bar{y}_2 - \bar{y}_1 = \delta_2 - \delta_1 + (\bar{x}_2 - \bar{x}_1)\beta + (\bar{u}_2 - \bar{u}_1)$. The repeat sales approach produces unbiased estimates if $\bar{x}_1 = \bar{x}_2$ and $\bar{u}_1 = \bar{u}_2$. Trivially, the mean values of x are the same over time if the sample is restricted to repeat sales. The means will also be the same over time for a matched sample if the matching procedure succeeds in balancing the period 1 and period 2 samples. The properties do not have to be the same over time as long as the average value of *x* does not change and $\bar{u}_1 = \bar{u}_2$.

The repeat sales approach's main weaknesses are that individual properties can change over time, and the coefficients in the underlying hedonic price function can change. Homes may be remodeled, and some neighborhoods may experience higher appreciation rates than others. The matching estimator suffers from similar problems. Although it may succeed in producing similar means over time for observable variables, there is no guarantee that it will produce identical means for omitted variables. An advantage of the repeat sales approach is that it guarantees that $\bar{x}_1 = \bar{x}_2$ for any variable that does change over time, whether the variable is observed or not. However, some of the apparent weaknesses of the matching approach may be overcome by using the observable explanatory variables to estimate the underlying hedonic price functions using the matched samples. Controlling for the observed X is directly comparable to standard hedonic estimation, which produces unbiased estimates if $cov(X, u) = 0$. As a result, it is generally preferable to use hedonic approaches to estimate price indices for matched samples.

8. Locally Weighted Quantile House Price Distribution Estimates

In this section, I present the results of locally weighted quantile regression estimates for the set of matched samples. It is important to recognize that the initial matching procedure did not include controls for location. The procedure produces annual samples that have approximately the same mean values for the observable structural variables. The approach does not necessarily produce the same number of sales from each neighborhood over time, however. Simple averages or medians may produce estimates of appreciation rates that are biased upward if the number of observations in later years happen to be drawn from high-priced neighborhoods. The repeat sales estimator may be subject to a similar bias if appreciation *rates* vary by neighborhood, but not if prices levels vary across neighborhoods while appreciation rates are the same everywhere. Thus, it is important to control for location when estimating price indices using a matching approach.

Controlling for location in the initial matching procedure is problematic. Including neighborhood indicator variables in the logit models could potentially produce samples with the same number of observations within each neighborhood over time. However, the number of observations is likely to be small unless all neighborhoods are large. Locally weighted versions of the logit models implicitly introduce each property's geographic coordinates as explanatory variables. When space is treated as continuous as is the case with locally weighted models, the only way to truly balance the samples is to limit them to repeat sales. Thus, controls for location need to be introduced at the second stage of the analysis – in the hedonic price functions.

I use the approach discussed in Section 4 to estimate quantile hedonic price regressions for the match samples. Equation (2) serves as the base equation. I estimate the models separately for Chicago, the North Suburbs, and the South Suburbs. The quantiles range from q = 0.03 to 0.97 at increments of 0.02, which implies B = 48 for each region. Taking advantages of the continuous structure of locally weighted procedures, I estimate the quantile regressions at a set of target locations and then interpolate to all other locations in the data set for each region. I use a tri-cube kernel with a 25% window based on straight-line distance between each observation and the target point.

The result of the estimation procedure is a set of $n \times B$ estimated sales prices for each region (n varies by region). The explanatory variables include the housing structural characteristics and the year of sale. Using the approach discussed in Section (4), I construct counterfactual density estimates by setting X to its actual values while setting D_t = 1 for the year under consideration and the values of the dummy variables to 0 for all other years. To simplify the presentation, I limit the results to three years – 2001, 2006, and 2011. This series of 5-year intervals is particularly interesting because it includes the time just before the start of the housing boom, the time near the peak, and a time well after the boom. To reduce the estimation time for such a large number of models, I omit data from other years when estimating the models.

The estimated sale price densities are shown in Figure 4 – 6. The sale price densities for 2001 are tightly clustered in all three regions, with relatively thin tails. The left tail of the distribution is larger in Chicago than in the suburbs in 2001. The distribution then shifts sharply to the right in all three regions in 2006. The rightward shift is larger for low sales prices in Chicago and the South Suburbs. The shift to the right for 2001 – 2006 is roughly the same at all price levels for the North Suburbs.

The collapse of the housing market after 2006 is clearly demonstrated by the pronounced leftward shifts in the sale price distributions between 2006 and 2011. The number of sales taking place at very low prices is even higher in 2011 than it was in 2001. The fat left tails are particularly evident in Chicago and the South Suburbs. The shift in the distribution for high-priced homes is much smaller. In all three regions, the rightward portions of the estimated densities for 2011 have returned to approximately their positions in 2006. Thus, the housing market collapse had a much greater effect on the low-priced portion of the market, particularly in Chicago and the South Suburbs.

To put the size of these shifts in perspective, it is useful to consider the implied shifts in the sale price densities when there are discrete changes in the values of other explanatory variables. Just as we can analyze the effect of a change in the year of sale by changing the values of the annual dummy variables while holding the structural characteristics at their actual values, we can simulate the effect of a change in a structural characteristic by comparing the estimated densities at two or more values of the variable while all other variables (including the year of sale) are set at their actual values. For any discrete variable j, I simply calculate kernel density estimates for the $n \times B$ estimated values of $\beta_j + \sum_{k \neq j} x_k \beta_k(q) + \sum_{t=2}^{T} D_t \delta_t(q)$ to the same expression with $\beta_j = 0$. The calculations for the continuous variables are similar, but the values for variable j are set to a set of representative values: building area is set to 1000, 2000, and 3000 square feet; lot size is set to 3000, 6000, and 9000 square feet; the number of rooms is 4, 6, or 8; the number of bathrooms is 1, 2, or 3; and the age is set to 25, 50, and 75. I pool the estimates across all three regions when estimating these density functions.

The estimated counterfactual densities are shown in Figure 7. Although Table 4 shows that all of these variables have significant effects in on sales prices, it is clear from Figure 7 that a change in any one of them has much less overall effect on the distribution of sales prices than the year of sale. The largest effects are for building area and lot size. As expected, increases in building and land areas shift the sale price distribution to the right. The effects of the two variables differ somewhat in their implications for the variance of the sale price distribution: whereas estimated sales prices are more variable for larger building areas, the variance declines as lot areas increase. There is a roughly parallel rightward shift in the sale price distribution as the number of bathrooms increases. The variance of the sale price distribution increases with housing age: whereas an increase in age has little effect on the density for high-priced homes, it shifts the distribution well to the left for low-priced homes.

The locally weighted quantile regressions allow all coefficients to vary smoothly over space. Although so far I have treated Chicago, the North Suburbs, and the South Suburbs as three separate regions, the estimates actually reveal a great deal of variation in appreciation rates within each of these areas. Figure 8 shows the estimated coefficients $\delta_{2006}(q)$ for q = 10%, 50%, and

90%. These results are directly comparable to the points shown for 2006 in Figure 2 for the three quantiles, but rather than having just one value for each regions, the estimated appreciation rates are allowed to vary smoothly over the region.

Figure 8 shows that the appreciation rate for 2000 – 2006 was exceptionally high in the South Side of the City of Chicago. The appreciation rate for the median was also relatively high in this area, but the region with high appreciation extends to the near West Side also. For q = 90%, the estimated appreciation rates are again high on the South and West Sides of the city, but there is also an area of high appreciation in the North Suburbs. The last panel of Figure 8 shows the difference between the appreciation rates at the 90% and the 10% quantile. The negative values for this difference on the city's South Side indicate that appreciation rates were much higher for the 10% quantile than for the 90% quantile. In contrast, in much of suburban Chicago – particularly the far North – appreciation rates were higher for the 90% quantile than for 10%. Overall, these results suggest that an assumption of a single housing market for the entire metro area is inappropriate.

Figure 9 shows that, just as prices rose most rapidly on the South and West Sides of Chicago during the boom, this is the same area where prices declined most markedly between 2006 and 2011. Rates of decline were much lower in the Near North area of Chicago and in much of the Northern and Western Suburbs. In all areas, prices declined much more at low prices than for high-priced homes.

9. Conclusion

The repeat sales estimator and hedonic price functions are the approaches most commonly used by economists to estimate house price indices. Both approaches focus on sample means. In fact, the repeat sales estimator is identical to period-by-period averages when the number of sales is the same in each period within the sample of repeat sales. Neither approach takes into account the possibility that rates of appreciation may be higher for low-priced homes during boom periods, with correspondingly great declines when the prices are falling.

In this paper, I show how quantile approaches can be used to analyze how the full distribution of quality-adjusted prices changes over time. I use local estimation procedures to allow appreciation rates to vary smoothly over space. The results show that rates of appreciation were indeed very high for low-priced homes in the South and West Sides of Chicago during the boom years of 2000 – 2006. Rates of decline were then also quite large in these areas. In contrast, prices for relatively high-priced homes in the relatively high-priced areas in Chicago Near North area and in the far Northern suburbs fell at much lower rates between 2006 and 2011 after having experienced substantial yet more moderate rates of appreciation between 2000 and 2006.

Another purpose of this paper has been to demonstrate how apparently complex locally weighted estimation of quantile models is feasible, while the large set of results remains easy to summarize graphically. Locally weighted quantile regressions produce separate predications for each observations and for each quantile. With 48 quantiles and more than 100,000 observations in each region, the models estimated here produce close to 500,000 sets of predicted values. Yet the results can be summarized easily using kernel density estimates to illustrate the counterfactual sale price distributions.

References

Archer, Wayne R., Dean H. Gatzlaff, and David C. Ling, "Measuring the Importance of Location in House Price Appreciation," *Journal of Urban Economics* 40 (1996), 334-353.

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

Bourassa, Steven C., Foort Hamelink, Martin Hoesli, and Bryan D. MacGregor, "Defining Housing Submarkets," *Journal of Housing Economics* 8 (1999), 160-183.

Bourassa, Steven C., Martin Hoesli, and Vincent S. Peng, "Do Housing Submarkets Really Matter?," *Journal of Housing Economics* 12 (2003), 12-28.

Carrillo, Paul E. and Jaren C. Pope, "Are Homes Hot or Cold Potatoes? The Distribution of Marketing Time in the Housing Market," *Regional Science and Urban Economics* 42 (2012), 189-197.

Case, Karl E. and Christopher J. Mayer, "Housing Price Dynamics within a Metropolitan Area," *Regional Science and Urban Economics* 26 (1996) 387-407.

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

Chaudhuri, Probal, "Nonparametric Estimates of Regression Quantiles and their Local Bahadur Representation," *Annals of Statistics* 19 (1991), 760-777.

Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica* (forthcoming).

Cleveland, William S. and Susan J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association* 83 (1988), 596-610.

Cobb-Clark, Deborah A. and Mathias G. Sinning, "Neighborhood Diversity and the Appreciation of Native- and Immigrant-Owned Homes," *Regional Science and Urban Economics* 41 (2011), 214-226.

Coulson, N. Edward and Daniel P. McMillen, "The Dynamics of Intra-Urban Quantile House Price Indexes," *Urban Studies* 44 (2007), 1517-1537.

Deng, Yongheng, Daniel McMillen, and Tien Foo Sing, "Private Residential Price Indices in Singapore: A Matching Approach," *Regional Science and Urban Economics* 42 (May 2012), 485-494.

Deng, Yongheng, Daniel McMillen, and Tien Foo Sing, "Matching Indices for Thinly-Traded Properties in Singapore," *Regional Science and Urban Economics* (forthcoming).

Goodman, Allen C. and Thomas G. Thibodeau, "Housing Market Segmentation," *Journal of Housing Economics* 7 (1998), 121-143.

Goodman, Allen C. and Thomas G. Thibodeau, "The Spatial Proximity of Metropolitan Area Housing Submarkets," *Real Estate Economics* 35 (2007), 209-232.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis* 15 (2007), 199-236.

Kostov, Philip. 2009. "A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices," *Spatial Economic Analysis*, 4, 53-72.

Liao, Wen-Chi and Xizhu Wang, "Hedonic House Prices and Spatial Quantile Regression," *Journal of Housing Economics* 21 (2012), 16-27.

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.

McMillen, Daniel P., "One Hundred Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics* 40 (1996), 100-124.

McMillen, Daniel P., "Neighborhood Price Indexes in Chicago: A Fourier Repeat Sales Approach," *Journal of Economic Geography* 3 (2003), 57-73.

McMillen, Daniel P., "Changes in the Distribution of House Prices over Time: Structural Characteristics, Neighborhood, or Coefficients?" *Journal of Urban Economics* 64 (2008), 573-589.

McMillen, Daniel P., "Repeat Sales as a Matching Estimator," *Real Estate Economics* 40 (2012a), 743-771.

McMillen, Daniel P., *Quantile Regression for Spatial Data*, Springer Briefs in Regional Science, New York (2012b).

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

Meese, Richard and Nancy Wallace, "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices," *AREUEA Journal* 19 (1991), 308-332.

Monkkonen, Paavo, Kelvin Wong, and Jaclene Begley, "Economic Restructuring, Urban Growth, and Short-Term Trading: The Spatial Dynamics of the Hong Kong Housing Market, 1992-2008," *Regional Science and Urban Economics* 42 (2012), 396-406.

Nicodemo, Catia and Josep Maria Raya, "Change in the Distribution of House Prices across Spanish Cities," *Regional Science and Urban Economics* 42 (2012), 739-748.

Ries, John and Tsur Somerville, "School Quality and Residential Property Values: Evidence from Vancouver Housing," *Review of Economics and Statistics* 92 (2010), 928-944.

Rouwendal, Jan and Simonetta Longhi, "The Effect of Consumers' Expectations in a Booming Housing Market: Space-Time Patterns in the Netherlands, 1999-2000," *Housing Studies* 23 (2008), 291-317.

Schmitz, Nick, Steven Shultz, and Roger Sindt, "Does A Rising Tide Raise All Ships? Single Family Housing Appreciation across Omaha (2000-2005), *Economics and Business Journal: Inquires and Perspectives* 1 (2008), 37-45.

Weber, Rachel, Saurav Dev Bhatta, and David Merriman, "Spillovers from Tax Increment Financing Districts: Implications for Housing Price Appreciation," *Regional Science and Urban Economics* 37 (2007), 259-281.

Yu, Keming and M. C. Jones., "Local Linear Quantile Regression," *Journal of the American Statistical Association* 93 (1998), 228-237.

Zahirovic-Herbert and Swarn Chatterjee, "Historic Preservation and Residential Property Values: Evidence from Quantile Regression," *Urban Studies* 49 (2012), 369-382.

Zeitz, Joachim, Emily Norman Zietz, and G. Stacy Sirmans. 2008. "Determinants of House Prices: A Quantile Regression Approach," *Journal of Real Estate Finance and Economics,* 37, 317-333.

Table 1

Number of Sales

Year	Cook County	Chicago	North Suburbs	South Suburbs
Full Sample				
2000	33,779	10,666	11,022	12,091
2001	39,315	12,809	12,626	13,880
2002	43,146	14,748	13,279	15,119
2003	45,605	15,354	14,384	15,867
2004	52,189	18,334	15,444	18,411
2005	52,487	18,852	14,667	18,968
2006	43,991	16,071	11,474	16,446
2007	31,050	11,140	8,464	11,446
2008	17,886	6,170	5,408	6,308
2009	16,609	5,856	5,135	5,618
2010	17,803	6,389	5,498	5,916
2011	16,084	5,695	5,133	5,256
Matched Sample				
2000	33,601	10,624	10,905	12,072
2001	33,598	10,623	10,903	12,072
2002	33,601	10,624	10,905	12,072
2003	33,593	10,623	10,898	12,072
2004	33,594	10,624	10,900	12,070
2005	33,601	10,624	10,905	12,072
2006	33,599	10,624	10,903	12,072
2007	30,506	10,624	8,458	11,424
2008	17,854	6,148	5,406	6,300
2009	16,572	5,830	5,129	5,613
2010	17,766	6,366	5,493	5,907
2011	16,043	5,676	5,121	5,246

Table 2

Descriptive Statistics

	Chicago		North Suburbs		South Suburbs	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Full Sample						
Log of Sale Price	12.0674	0.8098	12.6415	0.5743	12.0325	0.6557
Log of Building Area	7.0740	0.3214	7.3223	0.3857	7.1681	0.3458
Log of Lot Size	8.2238	0.3690	8.9035	0.6491	8.8160	0.5403
Rooms	5.5016	1.3968	6.3834	1.4464	5.9042	1.2607
Bathrooms	1.3521	0.5604	1.8090	0.7167	1.5195	0.6018
Brick	0.6100	0.4877	0.5777	0.4939	0.6674	0.4711
Basement	0.7668	0.4228	0.7848	0.4110	0.7777	0.4158
Central Air	0.2175	0.4126	0.5791	0.4937	0.4269	0.4946
Fireplace	0.1089	0.3115	0.4167	0.4930	0.2564	0.4366
Attic	0.4355	0.4958	0.2827	0.4503	0.3334	0.4714
Garage1	0.2224	0.4158	0.2140	0.4101	0.1452	0.3523
Garage2	0.5306	0.4991	0.6620	0.4730	0.7398	0.4387
Age	68.7264	28.5163	42.9575	20.5993	46.9617	23.5743
Number of Observations	142,084		122,534		145,326	
Matched Sample						
Log of Sale Price	12.0466	0.8258	12.6409	0.5816	12.0302	0.6698
Log of Building Area	7.0731	0.3157	7.3270	0.3851	7.1762	0.3480
Log of Lot Size	8.2358	0.3598	8.9001	0.6494	8.8165	0.5355
Rooms	5.4868	1.3709	6.4045	1.4472	5.9251	1.2700
Bathrooms	1.3342	0.5403	1.8134	0.7132	1.5269	0.6060
Brick	0.6192	0.4856	0.5746	0.4944	0.6805	0.4663
Basement	0.7717	0.4197	0.7833	0.4120	0.7876	0.4090
Central Air	0.2203	0.4144	0.5840	0.4929	0.4320	0.4954
Fireplace	0.1070	0.3092	0.4259	0.4945	0.2624	0.4399
Attic	0.4457	0.4970	0.2859	0.4518	0.3385	0.4732
Garage1	0.2261	0.4183	0.2140	0.4101	0.1491	0.3562
Garage2	0.5339	0.4989	0.6582	0.4743	0.7353	0.4412
Age	68.5244	27.9338	43.2743	20.7360	47.5568	23.7459
Number of Observations	109,010		105,926		118,992	

Table 3a

Means for Full and Matched Samples, Chicago

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Full Sample												
Log Price	11.81	11.87	11.96	12.04	12.12	12.26	12.32	12.33	12.17	11.85	11.80	11.71
Log Area	7.07	7.07	7.07	7.07	7.06	7.07	7.06	7.09	7.10	7.09	7.10	7.11
Lot Size	8.24	8.24	8.23	8.22	8.22	8.21	8.22	8.21	8.21	8.23	8.23	8.23
Rooms	5.46	5.46	5.47	5.49	5.48	5.48	5.48	5.57	5.60	5.57	5.57	5.63
Bathrooms	1.32	1.33	1.34	1.35	1.34	1.35	1.35	1.38	1.40	1.37	1.38	1.40
Brick	0.62	0.62	0.62	0.62	0.60	0.59	0.59	0.61	0.62	0.63	0.63	0.63
Basement	0.77	0.77	0.77	0.77	0.76	0.76	0.76	0.76	0.78	0.78	0.77	0.79
Central Air	0.22	0.22	0.23	0.23	0.20	0.21	0.20	0.22	0.23	0.23	0.24	0.24
Fireplace	0.10	0.10	0.11	0.11	0.10	0.10	0.10	0.12	0.13	0.12	0.13	0.13
Attic	0.45	0.45	0.44	0.43	0.44	0.43	0.43	0.42	0.43	0.44	0.43	0.43
Garage1	0.23	0.22	0.22	0.22	0.23	0.22	0.22	0.22	0.21	0.23	0.23	0.22
Garage2	0.53	0.53	0.53	0.53	0.52	0.53	0.52	0.53	0.54	0.54	0.53	0.55
Age	63.64	64.80	65.20	65.97	69.16	69.88	71.63	71.15	71.75	72.21	72.85	74.00
Matched Sample												
Log Price	11.81	11.87	11.95	12.04	12.15	12.28	12.34	12.30	12.16	11.85	11.79	11.71
Log Area	7.07	7.07	7.06	7.06	7.07	7.07	7.07	7.07	7.10	7.09	7.09	7.11
Log Lot size	8.24	8.24	8.24	8.24	8.24	8.24	8.24	8.23	8.21	8.23	8.23	8.23
Rooms	5.46	5.46	5.44	5.44	5.46	5.45	5.46	5.49	5.60	5.57	5.57	5.62
Bathrooms	1.32	1.32	1.31	1.31	1.32	1.32	1.32	1.34	1.40	1.37	1.38	1.40
Brick	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.61	0.62	0.62	0.63	0.63
Basement	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76	0.78	0.78	0.77	0.79
Central Air	0.22	0.22	0.21	0.21	0.22	0.22	0.22	0.21	0.23	0.23	0.23	0.24
Fireplace	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.13	0.12	0.13	0.13
Attic	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.43	0.44	0.44	0.43	0.43
Garage1	0.23	0.23	0.23	0.22	0.23	0.23	0.23	0.22	0.21	0.23	0.23	0.22
Garage2	0.53	0.53	0.54	0.53	0.53	0.53	0.53	0.53	0.54	0.54	0.53	0.55
Age	63.70	64.73	66.08	66.78	67.68	68.43	69.65	71.58	71.80	72.29	72.94	74.02

Table 3b

Means for Full and Matched Samples, North Suburbs

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Full Sample												
Log Price	12.35	12.43	12.54	12.61	12.70	12.80	12.86	12.88	12.80	12.63	12.61	12.50
Log Area	7.32	7.31	7.32	7.32	7.31	7.31	7.31	7.34	7.36	7.34	7.37	7.36
Lot Size	8.87	8.87	8.89	8.88	8.91	8.89	8.89	8.92	8.94	8.94	8.99	8.98
Rooms	6.37	6.33	6.38	6.36	6.34	6.35	6.33	6.44	6.51	6.45	6.53	6.51
Bathrooms	1.80	1.79	1.81	1.81	1.80	1.79	1.78	1.83	1.87	1.83	1.87	1.86
Brick	0.56	0.57	0.57	0.57	0.57	0.58	0.57	0.57	0.60	0.61	0.62	0.62
Basement	0.77	0.77	0.78	0.78	0.78	0.78	0.78	0.79	0.80	0.80	0.83	0.82
Central Air	0.58	0.59	0.58	0.59	0.58	0.57	0.57	0.57	0.60	0.57	0.58	0.58
Fireplace	0.42	0.40	0.41	0.41	0.41	0.40	0.40	0.43	0.46	0.43	0.46	0.46
Attic	0.28	0.27	0.27	0.27	0.28	0.28	0.29	0.29	0.30	0.29	0.30	0.31
Garage1	0.22	0.22	0.22	0.21	0.22	0.22	0.21	0.21	0.21	0.21	0.20	0.20
Garage2	0.65	0.66	0.66	0.66	0.66	0.65	0.65	0.67	0.68	0.68	0.69	0.70
Age	38.77	38.50	39.50	40.76	42.36	43.22	45.04	45.99	47.00	48.38	50.46	51.70
Matched Sample												
Log Price	7.32	7.32	7.32	7.32	7.32	7.32	7.32	7.34	7.36	7.34	7.37	7.36
Log Area	8.88	8.88	8.88	8.88	8.88	8.88	8.89	8.92	8.94	8.94	8.99	8.98
Log Lot size	6.38	6.37	6.37	6.36	6.38	6.38	6.38	6.44	6.51	6.45	6.53	6.51
Rooms	1.80	1.80	1.79	1.80	1.80	1.80	1.80	1.83	1.87	1.83	1.86	1.86
Bathrooms	0.56	0.56	0.56	0.56	0.56	0.56	0.57	0.57	0.60	0.61	0.61	0.62
Brick	0.77	0.78	0.77	0.78	0.77	0.77	0.78	0.79	0.80	0.80	0.83	0.83
Basement	0.59	0.59	0.58	0.58	0.59	0.58	0.59	0.58	0.60	0.57	0.58	0.58
Central Air	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.43	0.46	0.43	0.46	0.46
Fireplace	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.30	0.29	0.30	0.31
Attic	0.22	0.22	0.22	0.22	0.21	0.22	0.22	0.21	0.21	0.21	0.20	0.20
Garage1	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.67	0.68	0.68	0.69	0.70
Garage2	38.36	39.15	40.31	41.44	42.14	43.24	44.39	45.97	47.00	48.36	50.44	51.68
Age	7.32	7.32	7.32	7.32	7.32	7.32	7.32	7.34	7.36	7.34	7.37	7.36

Table 3c

Means for Full and Matched Samples, South Suburbs

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Full Sample												
Log Price	11.84	11.88	11.94	12.03	12.09	12.18	12.22	12.22	12.11	11.88	11.83	11.68
Log Area	7.17	7.17	7.17	7.17	7.16	7.15	7.15	7.17	7.19	7.19	7.20	7.21
Lot Size	8.81	8.81	8.81	8.81	8.82	8.81	8.81	8.82	8.84	8.83	8.84	8.85
Rooms	5.91	5.89	5.91	5.92	5.89	5.86	5.85	5.88	5.98	5.96	6.01	6.02
Bathrooms	1.52	1.52	1.52	1.52	1.51	1.51	1.50	1.51	1.55	1.53	1.56	1.57
Brick	0.69	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.68	0.68	0.68	0.68
Basement	0.79	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.78	0.80	0.80	0.80
Central Air	0.43	0.44	0.43	0.43	0.42	0.41	0.41	0.42	0.44	0.43	0.45	0.43
Fireplace	0.26	0.25	0.26	0.26	0.25	0.24	0.24	0.25	0.28	0.27	0.29	0.28
Attic	0.34	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.34
Garage1	0.15	0.14	0.14	0.15	0.15	0.14	0.14	0.15	0.14	0.14	0.14	0.14
Garage2	0.73	0.75	0.74	0.74	0.74	0.74	0.74	0.73	0.74	0.75	0.75	0.76
Age	43.15	43.72	44.35	45.22	46.53	47.32	48.28	49.30	49.51	51.30	52.67	53.99
Matched Sample												
Log Price	11.84	11.89	11.95	12.04	12.11	12.21	12.26	12.22	12.11	11.88	11.83	11.68
Log Area	7.17	7.17	7.17	7.17	7.17	7.17	7.17	7.17	7.19	7.19	7.20	7.21
Log Lot size	8.81	8.81	8.81	8.81	8.81	8.81	8.81	8.82	8.84	8.83	8.84	8.85
Rooms	5.91	5.92	5.92	5.92	5.90	5.91	5.91	5.88	5.98	5.96	6.01	6.02
Bathrooms	1.52	1.52	1.52	1.53	1.52	1.52	1.52	1.51	1.55	1.54	1.56	1.57
Brick	0.69	0.68	0.68	0.68	0.68	0.69	0.68	0.65	0.68	0.68	0.68	0.68
Basement	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.77	0.78	0.80	0.80	0.80
Central Air	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.42	0.44	0.43	0.45	0.43
Fireplace	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.28	0.27	0.29	0.28
Attic	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.33	0.33	0.33	0.33	0.34
Garage1	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.14	0.14	0.14	0.14
Garage2	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.74	0.75	0.75	0.76
Age	43.14	44.24	45.23	46.18	47.39	47.98	49.07	49.27	49.50	51.30	52.67	53.99

Table 4

Regression Results for Log of Sale Price

Variable	Chicago		North Suburbs		South Suburbs	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Log of Building Area	0.2943	0.0075	0.3554	0.0056	0.3325	0.0060
Log of Lot Size	0.3135	0.0059	0.2081	0.0020	0.1796	0.0027
Rooms	-0.0036	0.0017	0.0067	0.0012	0.0147	0.0014
Bathrooms	0.0398	0.0039	0.0510	0.0022	0.0511	0.0027
Brick	0.0518	0.0040	-0.0106	0.0024	0.0343	0.0028
Basement	0.0042	0.0046	0.0962	0.0026	0.1692	0.0030
Central Air	0.0066	0.0041	0.0089	0.0022	0.0200	0.0025
Fireplace	0.0379	0.0056	0.0404	0.0024	0.0500	0.0031
Attic	-0.0092	0.0034	0.0117	0.0022	0.0070	0.0025
Garage1	0.0404	0.0043	0.0230	0.0035	0.0468	0.0042
Garage2	0.0560	0.0038	0.0632	0.0033	0.0768	0.0035
Age	-0.0024	0.0001	-0.0014	0.0001	-0.0030	0.0001
2001 Sale	0.0762	0.0064	0.0855	0.0041	0.0557	0.0047
2002 Sale	0.1765	0.0064	0.1631	0.0041	0.1356	0.0047
2003 Sale	0.2962	0.0064	0.2410	0.0041	0.2273	0.0047
2004 Sale	0.4247	0.0064	0.3324	0.0041	0.3174	0.0047
2005 Sale	0.5799	0.0064	0.4420	0.0041	0.4345	0.0047
2006 Sale	0.6654	0.0064	0.4968	0.0041	0.5074	0.0047
2007 Sale	0.6286	0.0064	0.4812	0.0044	0.4880	0.0048
2008 Sale	0.3866	0.0075	0.3614	0.0050	0.3052	0.0057
2009 Sale	0.0493	0.0076	0.2084	0.0051	0.0539	0.0059
2010 Sale	-0.0113	0.0074	0.1346	0.0050	-0.0076	0.0058
2011 Sale	-0.1284	0.0077	0.0343	0.0052	-0.1675	0.0060
Constant	7.1414	0.0618	7.7125	0.0372	7.5678	0.0414
Number of Census Tract Fixed Effects	807		225		268	
Number of Observations	109,010		105,926		118,992	
R^2	0.6863		0.7338		0.7098	

Figure 1

Estimated Hedonic Price Indices

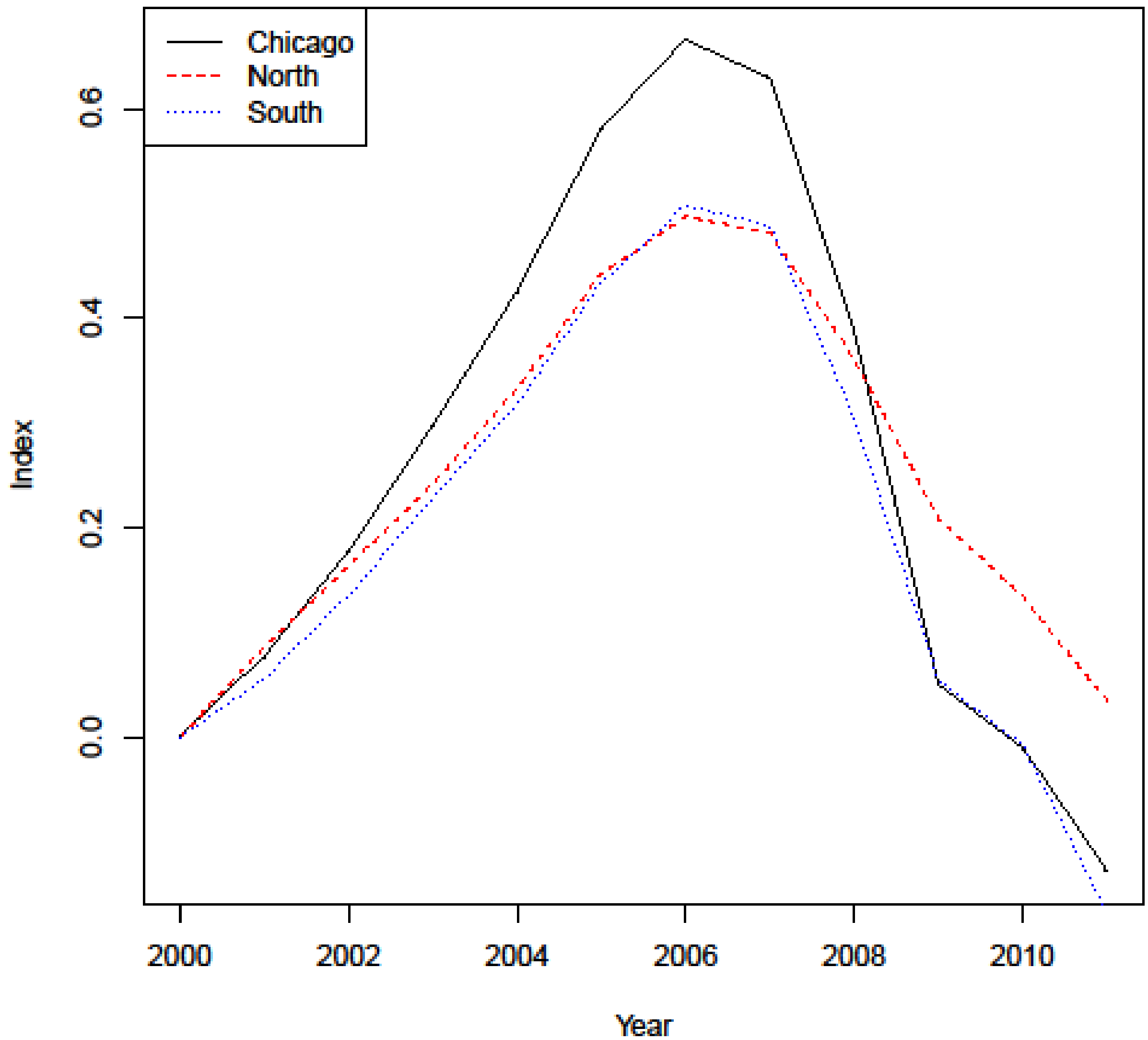

Figure 2

Sale Price Percentiles (10%, 50%, 90%) by Year

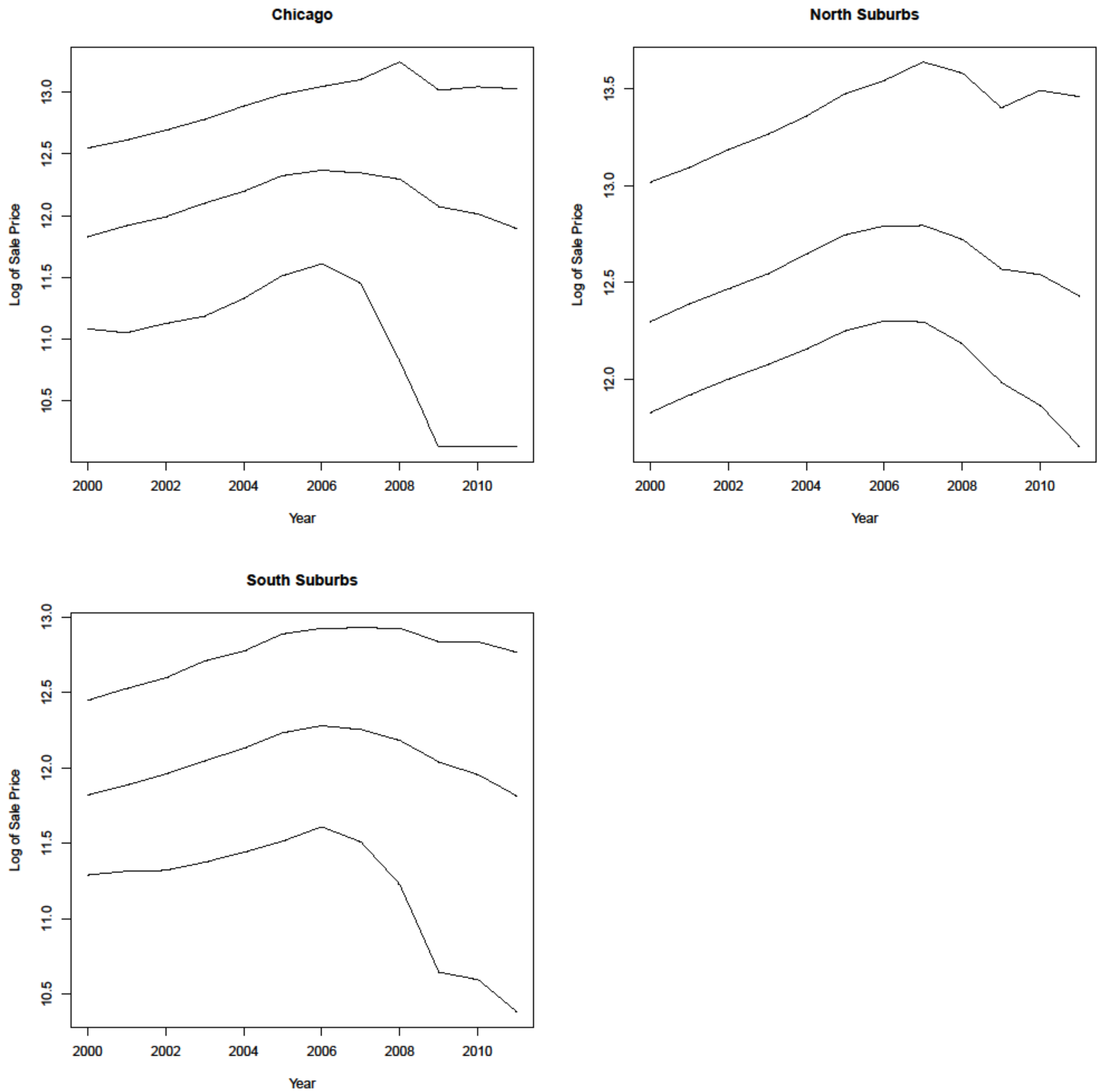

Figure 3

Sale Price Percentile Indices

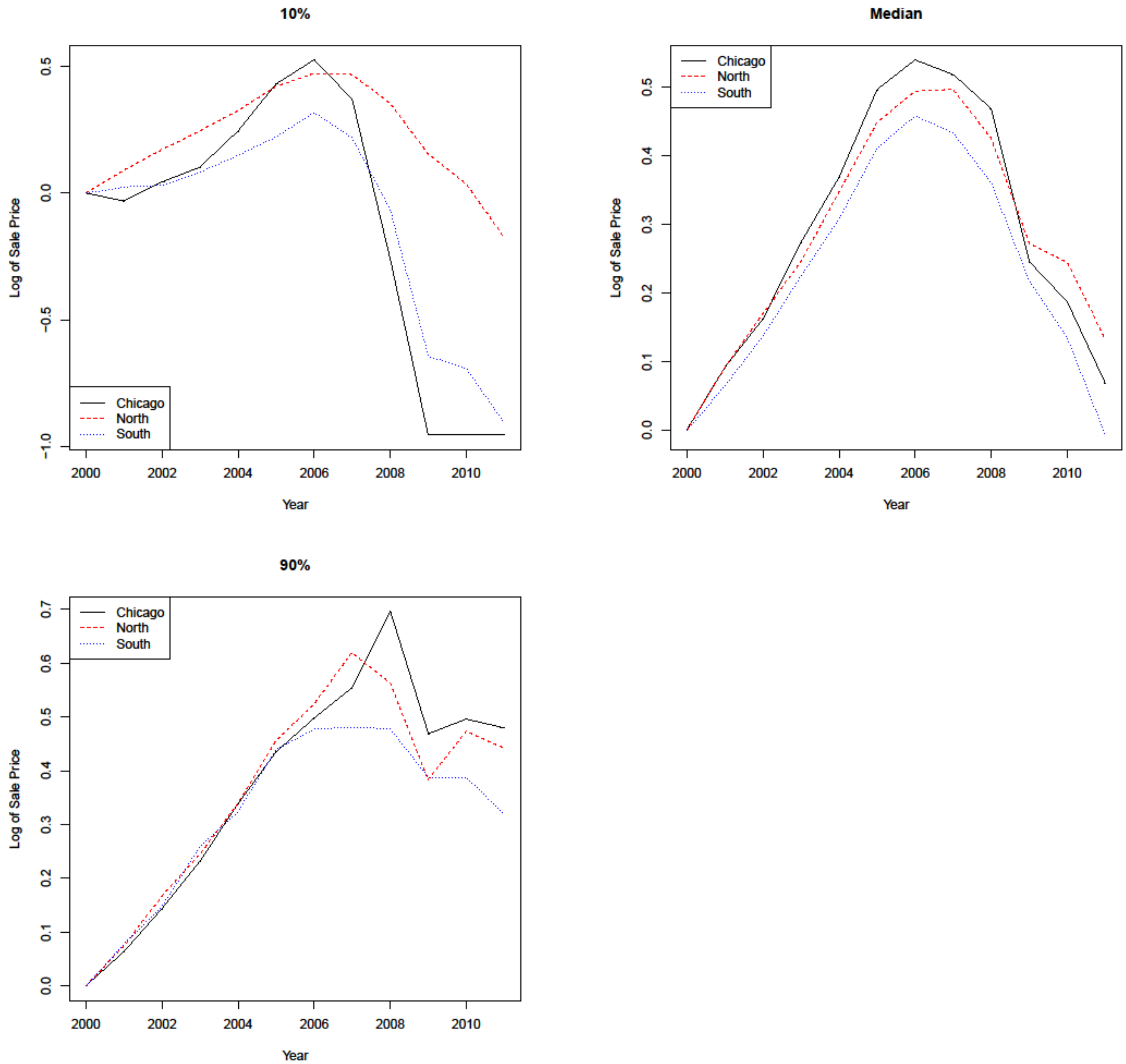

Figure 4

Estimated Sale Price Densities for Chicago

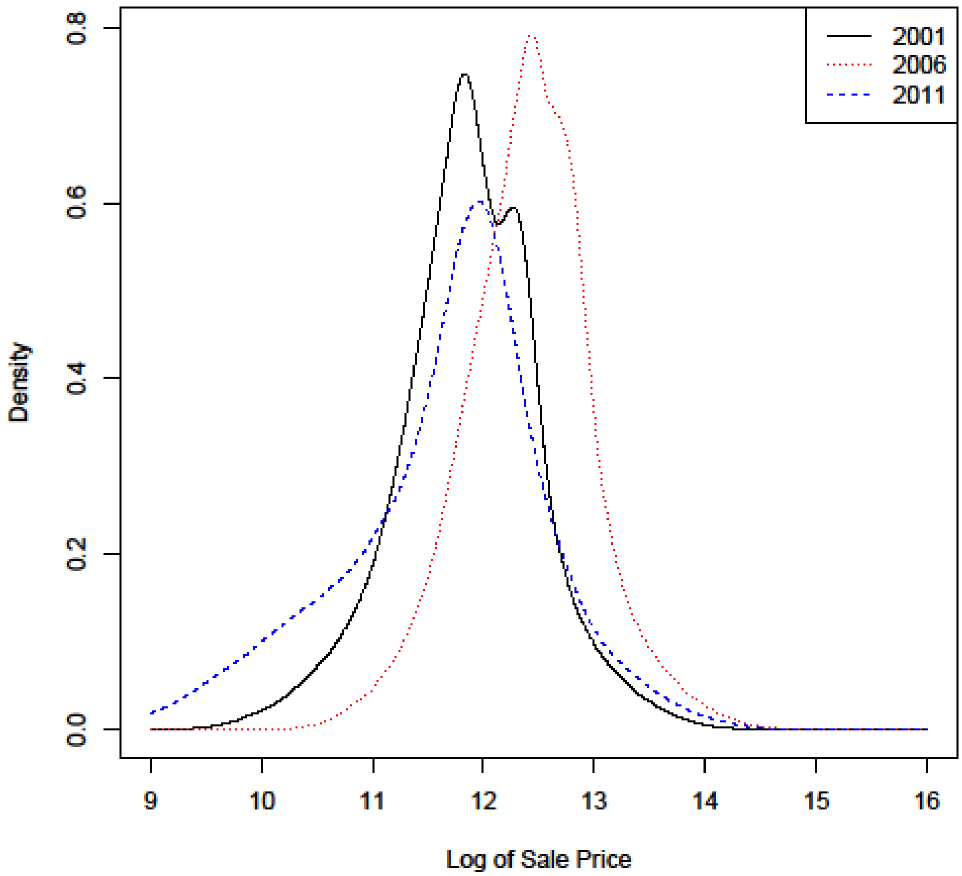

Figure 5

Estimated Sale Price Densities for North Suburbs

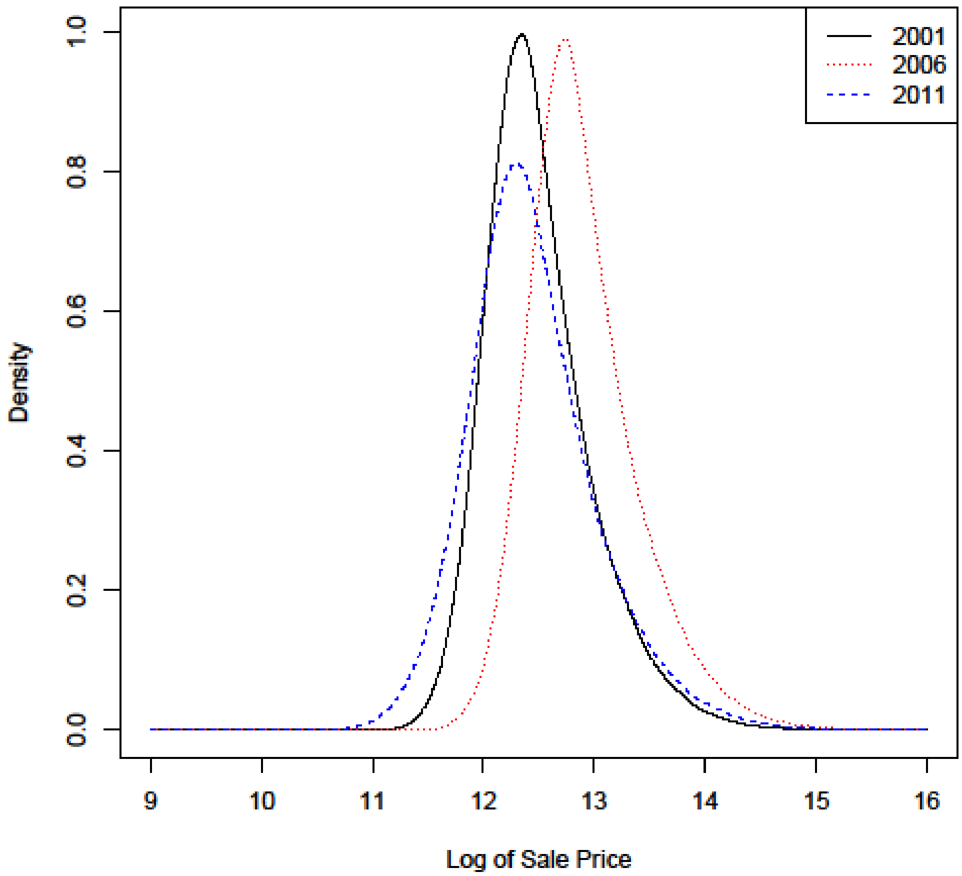

Figure 6

Estimated Sale Price Densities for South Suburbs

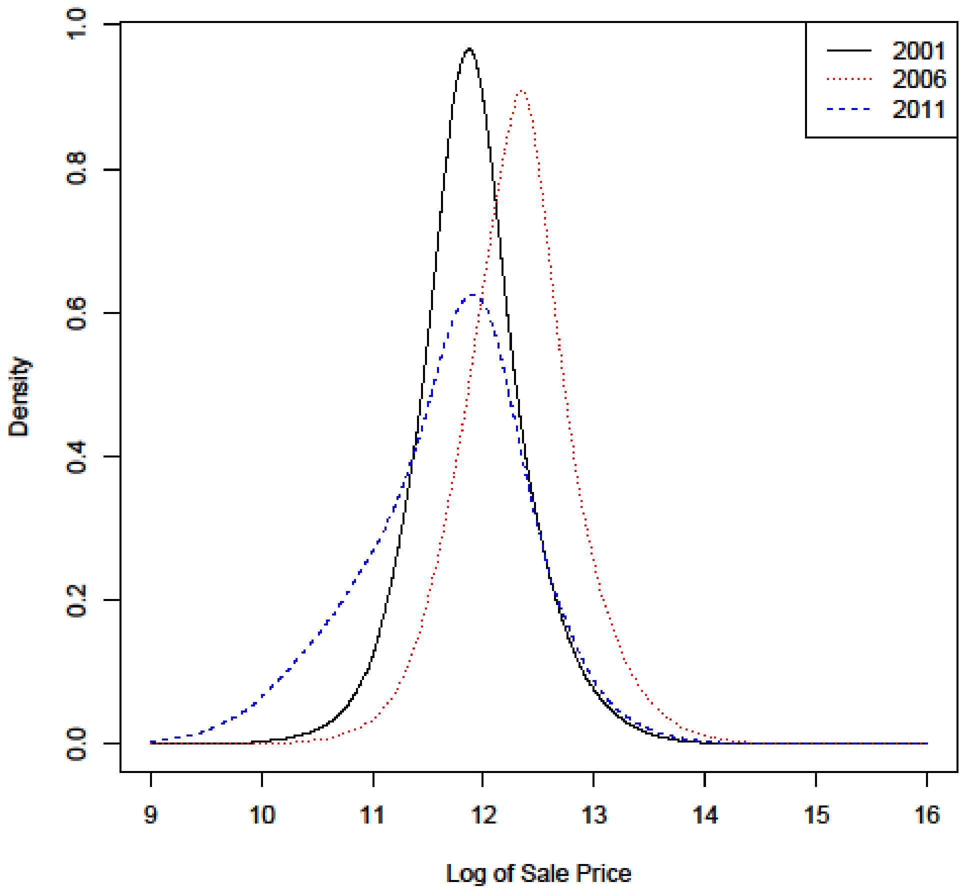

Figure 7

Partial Effects of Changes in Housing Characteristics

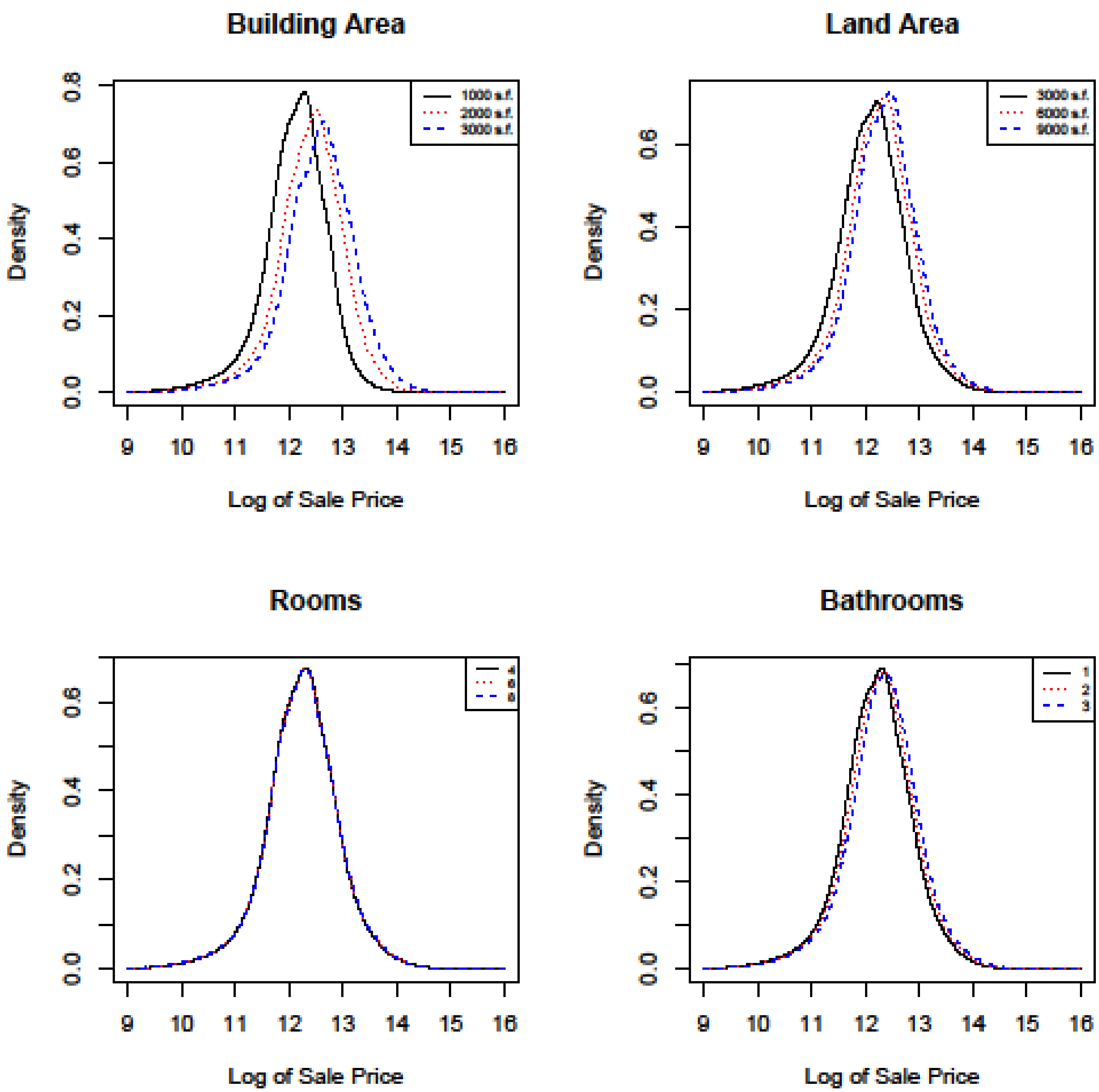

Figure 7 (cont'd)

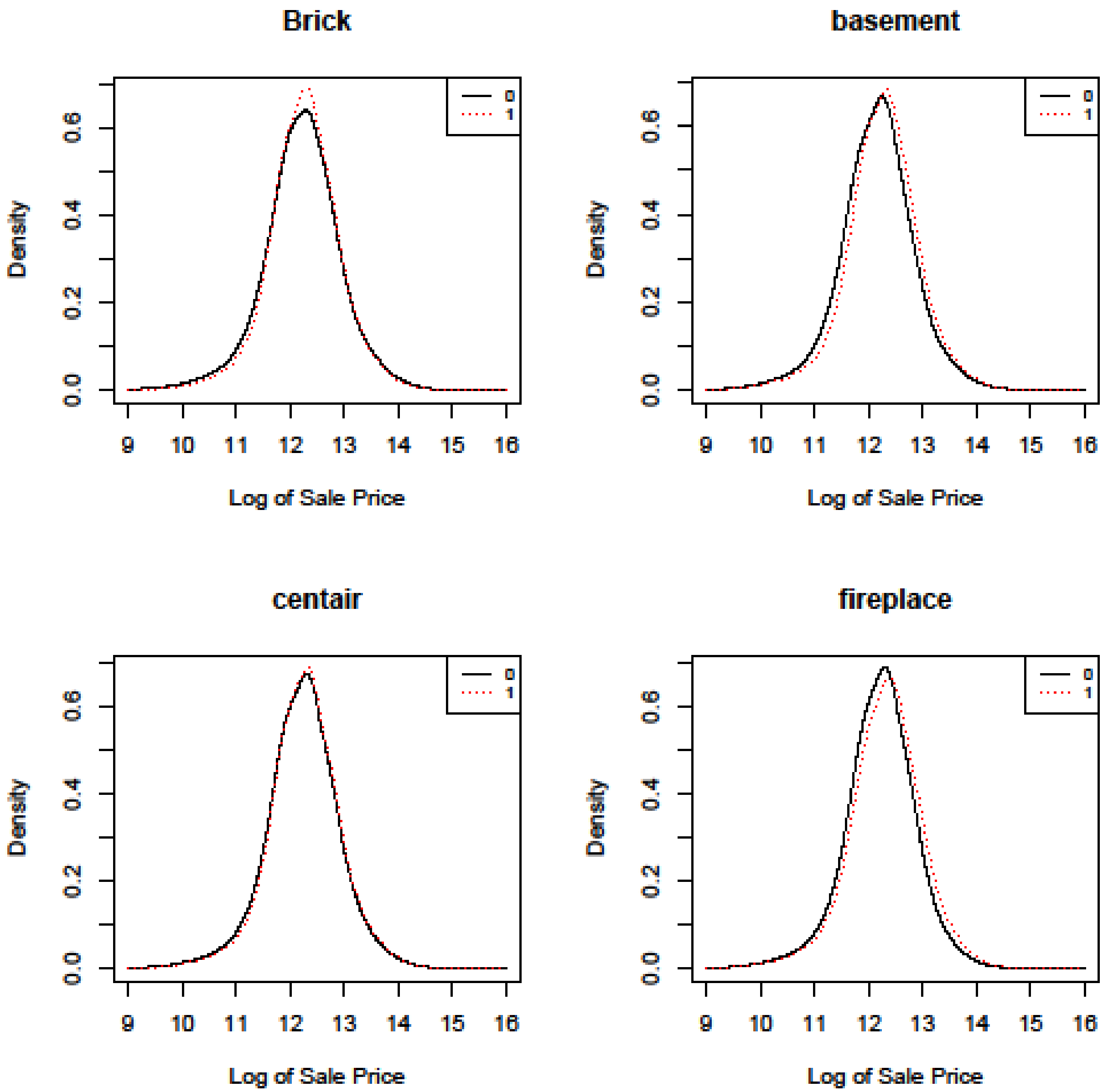

Figure 7 (cont'd)

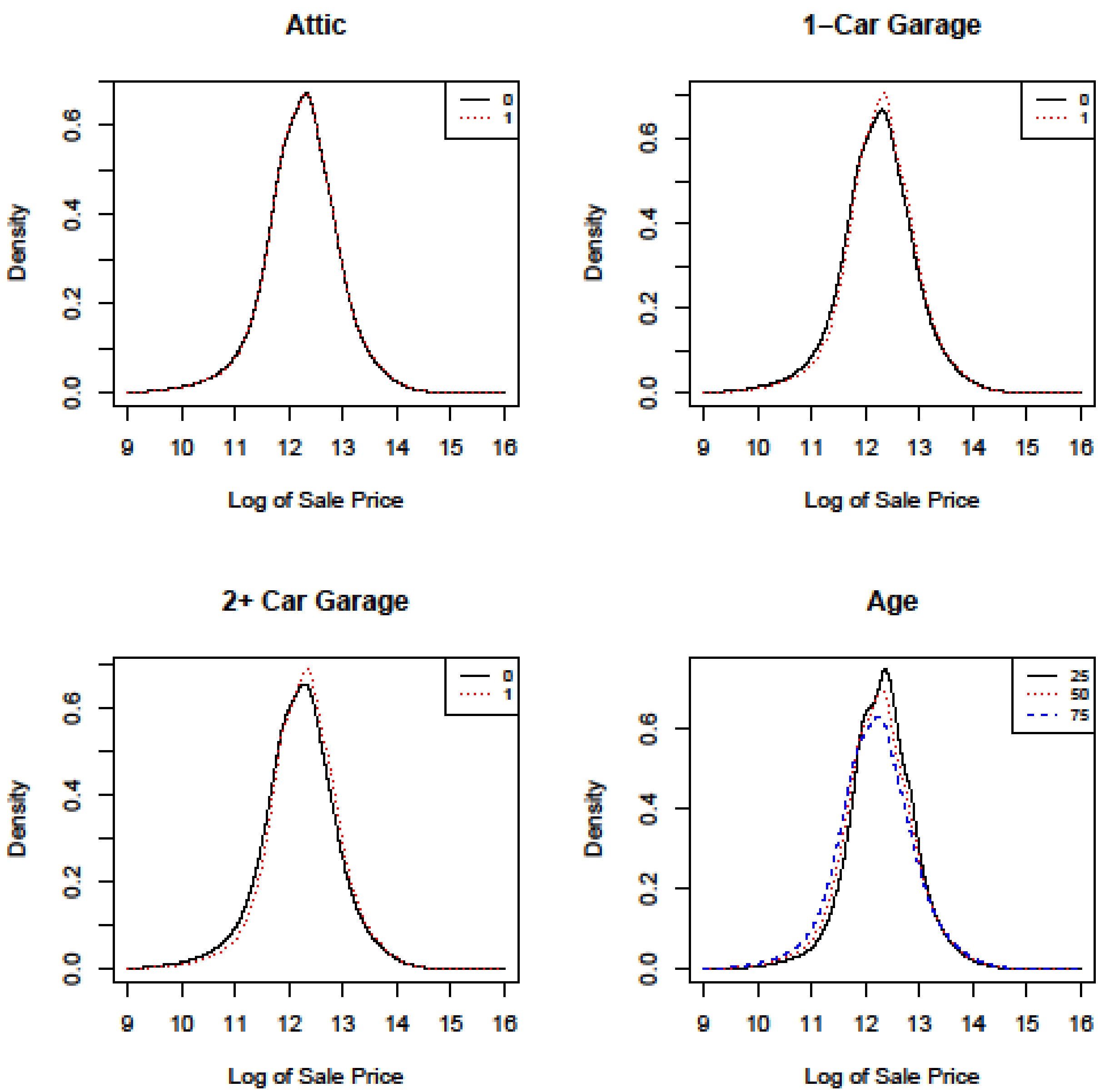

Figure 8

Estimated Appreciation Rates, 2000 - 2006

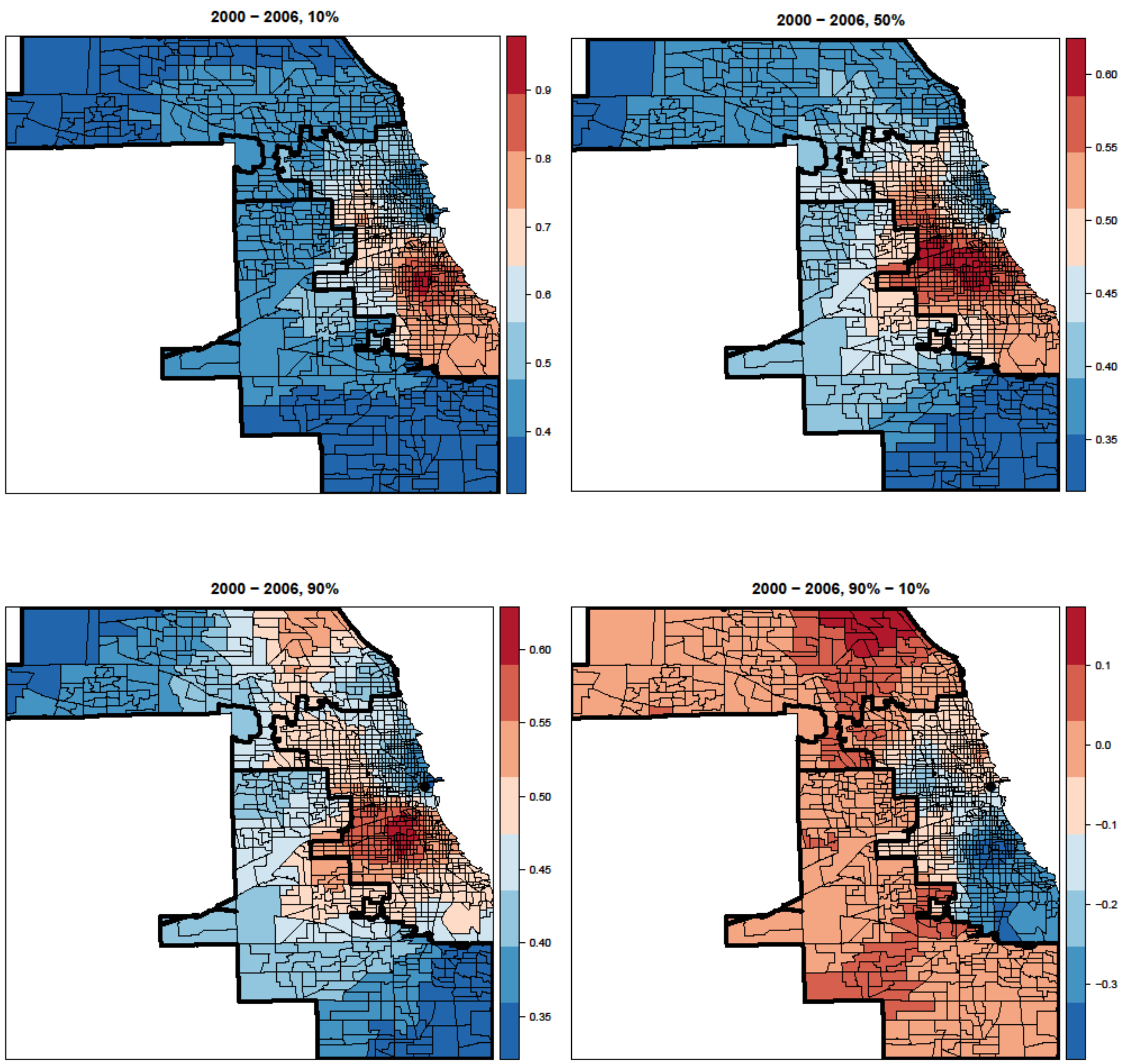

Figure 9

Estimated Appreciation Rates, 2006 - 2011