Grant-in-Aid for Scientific Research（S）

Real Estate Markets, Financial Crisis, and Economic Growth

: An Integrated Economic Approach

Working Paper Series No.74

Decompositions of Spatially Varying Quantile Distribution Estimates: The Rise and Fall of Tokyo House Prices

Daniel McMillen
and
Chihiro Shimizu

December, 2017

HIT-REFINED PROJECT
Institute of Economic Research, Hitotsubashi University
Naka 2-1, Kunitachi-city, Tokyo 186-8603, JAPAN
Tel: +81-42-580-9145
E-mail: hit-refined-sec@ier.hit-u.ac.jp
http://www.ier.hit-u.ac.jp/hit-refined/

Decompositions of Spatially Varying Quantile Distribution Estimates: The Rise and Fall of Tokyo House Prices*

Daniel McMillen
Department of Economics
University of Illinois
1407 W. Gregory Dr.
Urbana, IL 61801
mcmillen@illinois.edu

Chihiro Shimizu
Nihon University, Setagaya Campus
3-34-1 Shimouma, Setagaya-ku
Tokyo 154-8513 Japan
shimizu.chihiro@nihon-u.ac.jp

This draft: Dec 2017

Abstract

We extend Machado-Mata's (2005) approach for decomposing the differences in the distribution of a dependent variable across two samples to account for location when the models are estimated using conditional parametric procedures. We find that a substantial portion of the change in the distribution of condominium prices in Tokyo between the rapid rise in prices in 1986 – 1990 and the sharp decline in 1991 – 1995 is due to changes in the values of the explanatory variables. Changes in the locations of sales serve to shift the price distribution to the left because later sales were more likely to be farther from downtown Tokyo, where prices are lower.

Keywords: Conditionally parametric, quantile regression, decomposition.

JEL codes: C14, C18, R30

* We gratefully acknowledge the financial support of JSPS KAKEN(S) #25220502, Nomura Foundation.

1. Introduction

Although quantile regression is designed to estimate the conditional distribution of a dependent variable given the values of the explanatory variables, conditional quantile estimates also imply a full marginal distribution when the regressions are estimated at multiple quantiles. In an influential paper, Machado and Mata (2005) show how quantile regression estimates can be used to estimate counterfactual marginal distributions. Estimating quantile regressions at many quantiles implies a predicted distribution of the dependent variable at given values of the explanatory variables. The parametric structure of the conditional quantile model allows the differences in the predicted distributions across two samples to be decomposed into portions that are due to changes in the explanatory variables and changes in the explanatory variables.

Two approaches have been proposed for estimating quantile regression models for spatial data. A fully parametric approach typically uses the simple spatial autoregressive model, $Y = \rho WY + X\beta + u$ as the base estimating equation, where W is an $n \times n$ weight matrix showing the influence of each value of y_j on y_i. Quantile regression versions of the model can be estimated using the instrumental variable approaches of Chernozhukov and Hansen (2006) or Kim and Muller (2004), which account for the endogeneity of WY. Examples of these approaches include Kostov (2009), Lei and Zhang (2017), Liao and Wang (2012), Zeitz et al. (2008), and Zhang and Leonard (2014). An alternative is to estimate a "geographically weighted regression" version of the quantile regression model. This approach, which was proposed by McMillen (2013, 2016), is based on the conditionally parametric (CPAR) approach of Cleveland, Grosse, and Shyu (1992) and Cleveland (1994). The geographic version of the CPAR approach estimates separate coefficients for various target sets of geographic coordinates. Conditional on location, the model is a standard quantile regression model, but the coefficients vary smoothly over space. The CPAR approach is most useful when a base parametric specific is appropriate for small geographic regions but the base model does not hold over the entire region covered by the data.

When separate CPAR models are estimated for two samples, the underlying estimating equations can be written as $y_1 = x_1'\beta_1(z_1) + u_1$ and $y_2 = x_2'\beta_2(z_2) + u_2$ where z_1 and z_2 are the vectors of geographic coordinates (i.e., longitude and latitude) for samples 1 and 2. Thus, differences in the distribution of y across the two samples are attributable to three sources – differences in the explanatory variables (x), differences in the coefficients (β), and differences in the locations of the observations. In this paper, we show that it is straightforward to extend the Machado-Mata (2005) approach for decomposing the differences in the distribution of y across two samples to also account for differences in the location of the observations.

Our data set includes all sales of condominiums in Tokyo from 1986 – 2016. We focus on a particularly interesting time from 1986 – 1995. Prices more than doubled from 1986 to the end of 1990, after which they declined by more than 70% from the beginning of 1991 to the end of 1995. Although the rate of decline moderated afterward, the decline in prices continued through the end of 2001, and prices were still not back to their peak levels at the end of 2016. The 1986 – 1995 period includes the periods of rapid rise and subsequent rapid decline in prices.

We use the Machado-Mata (2005) approach and our extension to decompose the change in the distribution of prices between the rapid rise for 1986 – 1990 and the rapid decline for 1991 – 1995 into the portions that are due to changes in the estimated quantile coefficients, changes in the values of the explanatory variables, and changes in the locations of the sales. The primary explanatory variables are the area and age of the unit, the story, and whether the unit has a southern exposure. In addition, we include controls for location and the quarter of sale. Focusing on the primary explanatory variables, we find that a large portion of the change in the distributions is attributable to changes in the values of the variables – primarily the age of the units, which naturally tends to become higher over time despite a good amount of new construction. The change in the locations of sales between 1986 – 1990 and 1991-1995 serves to shift the distribution of sales prices further to the left because sales in the later period were more likely to be in locations farther from downtown Tokyo, where prices are lower.

2. Spatial Quantile Regression

The standard quantile regression estimating equation can be written as $Q_y(\tau|x_i) = x_i'\beta$. This equation implies that the conditional quantile function for the dependent variable y at quantile τ is a linear function of x_i. Following Koenker (2005), the standard quantile regression approach involves finding the values for $\hat{\beta}(\tau)$ that minimize $\sum_i \rho_\tau(y_i - x_i'\beta)$, where $\rho_\tau(u)$ is the piecewise linear function $\rho_\tau(u) = u\big(\tau - I(u < 0)\big)$. Nonparametric approaches for the quantile regression model can be implemented by adding a kernel weight function to this expression. At the target, x, the objective function is $\sum_i \rho_\tau w_i(x)(y_i - (x_i - x)'\beta)$, where $w_i(x) = K\big((x_i - x)/h\big)/h$ is the kernel weight function and h is the bandwidth (Chaudhuri, 1991; Koenker, 2005, chapter 7; Yu and Jones, 1998).

The conditionally parametric approach differs from this fully nonparametric locally weighted quantile regression approach by having different variables in the kernel weight function and the base regression:

$$\min_\beta \sum_{i=1}^{n} w_i(z)\rho_\tau(y_i - x_i'\beta) \quad (1)$$

For spatial models, z typically represents either longitude and latitude or the straight-line distance between observation i and the target location. A smaller bandwidth leads to more local variation in the estimated coefficients, $\hat{\beta}(\tau, z)$, which vary by both quantile and location.[2] The approach is discussed in more detail in McMillen (2013, 2015), who used an adaptive decision tree approach (Loader, 1999, section 12.2) to choose the set of target points, after which the results are interpolated to every point in the data set. The adaptive decision tree approach chooses more locations in areas with many observations, and smaller bandwidths lead to more target points. In this application, we choose instead to divide the sample area into a set of square kilometer grid cells, and then interpolate from the cell midpoints to the full set of locations represented in the data set. The advantage of this approach is that it directly corresponds to the regions covered in the maps we use to summarize the results.[3]

After interpolation, there are n coefficient estimates for each value of τ, one for each observation. McMillen (2013, 2015) shows how a procedure based on Machado-Mata (2005) can be used to estimate the marginal distribution of y implied by the CPAR conditional quantile estimates. At a given quantile τ, the quantile regression prediction for observation i is simply $x_i'\hat{\beta}(\tau, z_i)$. In matrix form, the full set of n estimates is $X \circ \hat{\beta}(\tau, z) e_k$, where X and $\hat{\beta}(\tau, z)$ are both $n \times k$ matrices, $e_k = (1 \quad \ldots \quad 1)$, and $\circ$ represents the Hadamard product. The predictions across T values of τ can then be stored in the $n \times T$ matrix

[2] In the empirical section of the paper, we use a tri-cube kernel with a 25% window, which means that the nearest 25% of the observations receive weight when estimating the coefficients at a target location. Weights are based on straight-line distances, i.e., $w_i(z) = I(d_i \le d_{.25}) \left(1 - \left(\frac{d_i}{d_{.25}}\right)^3\right)^3 / d_{.25}$, where d_i is the distance between observation i and the target location $d_{.25}$ is the 25[th] percentile of the distances.

[3] The adaptive decision tree approach is much less computation intensive, producing 69 target points for 1986 – 1990 and 71 target points for 1991 – 1990, compared with 845 square-kilometer grid cells.

$\hat{Y} = \{X \circ \hat{\beta}(\tau_1, z)e_k \quad \ldots \quad X \circ \hat{\beta}(\tau_T, z)e_k\}$. The implied distribution of $\hat{Y}$ can then be calculated using a standard kernel density estimator for the implied nT-vector of predictions.

Kernel density estimates can also be used to show how the distribution of y responds to discrete changes in an explanatory variable. For example, consider a two-explanatory variable model in which $X = (1 \quad x_1 \quad x_2)'$ and $\beta = (\beta_0 \quad \beta_1 \quad \beta_2)'$. Setting x_1 to an arbitrary value δ, the predicted values at a given quantile are $\hat{\beta}_0 + \delta\hat{\beta}_1 + x_2 \circ \hat{\beta}_2$, where $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are each n-vectors. After grouping these estimates in to the $n \times T$ matrix $\hat{y}(\delta)$, a kernel density estimator can be used to calculate the implied distribution of $\hat{y}$ when $x_1 = \delta$ and x_2 is set to its actual set of values in the data set. This procedure can be repeated at various values of δ to show how the distribution of predicted values of y changes as x_1 changes.

3. Counterfactual Decompositions of Distribution Changes

Machado and Mata (2005) present a simple extension of the Oaxaca (1973) approach for decomposing the difference in two sets of estimates to the portions due to differences in the estimated coefficients and the values of the explanatory variables. For linear regression estimates, the Oaxaca (1973) decomposition is $\hat{y}_1 - \hat{y}_2 = x_1'\hat{\beta}_1 - x_2'\hat{\beta}_2 = \left(x_1'\hat{\beta}_1 - x_2'\hat{\beta}_1\right) + \left(x_2'\hat{\beta}_1 - x_2'\hat{\beta}_2\right)$. The expression in the first set of parentheses shows the effect of differences in the explanatory variables on the differences in predicted values, while the terms in the second set of parentheses represent the effect of differences in the coefficients. These expressions are typically evaluated at the mean values of the explanatory variables, so the fact that the number of observations will usually differ across the two samples does not affect the calculations. The order of the decomposition can be changed so that the change in coefficients precedes the change in variables.

Machado and Mata's (2005) version of the decomposition shows the effect of differences in the variables and estimated quantile regression coefficients on differences in the full distribution of predicted values of the dependent variables.[4] They propose sampling with replacement from the rows of the explanatory variables X_1 and X_2 to form the $M \times k$ matrices X_1^m and X_2^m, where M is the number of draws. They also propose re-estimating the quantile regressions at sets of randomly drawn values of τ. A computationally less burdensome approach is to estimate the quantile regressions at a series of fixed values of τ, such as τ = 0.02, 0.03, … , 0.98, which produces two $k \times T$ matrices of coefficients, $\hat{\beta}_1^q$ and $\hat{\beta}_2^q$, where the superscript indicates that the matrices are coefficient vectors estimated at various quantile values. The matrices of predicted values needed for the decompositions are simply $X_1^m\hat{\beta}_1^q$, $X_2^m\hat{\beta}_1^q$, and $X_2^m\hat{\beta}_2^q$, all of which are of dimension $M \times T$, i.e., M draws from the rows of the X matrices by T quantiles. Each of these terms can then be treated as MT-vectors, and kernel density estimates can be calculated for each vector.

This approach extends readily to the case of spatial CPAR quantile regression. Following Machado and Mata's (2005) approach, let X_1^m and X_2^m represent matrices of constructed by making M draws with replacement from the rows of X_1 and X_2. Similarly, for each value of τ, construct the $M \times k$ matrices $\hat{\beta}_1^m(\tau, z_1)$ and $\hat{\beta}_2^m(\tau, z_2)$ by drawing randomly with replacement from the rows of $\hat{\beta}_1(\tau, z_1)$ and $\hat{\beta}_2(\tau, z_2)$. The predicted values needed for the decompositions then are constructed using equations (2) – (4):

---

[4] Examples of studies using the Machado and Mata (2005) approach include Carrillo and Yezer (2009) and McMillen (2008).

$$\hat{Y}_1^m = \{X_1^{m\circ}\hat{\beta}_1^m(\tau_1, z_1)e_k \quad \dots \quad X_1^{m\circ}\hat{\beta}_1^m(\tau_T, z_1)e_k\} \tag{2}$$

$$\hat{Y}_2^m = \{X_2^{m\circ}\hat{\beta}_2^m(\tau_1, z_2)e_k \quad \dots \quad X_2^{m\circ}\hat{\beta}_2^m(\tau_T, z_2)e_k\} \tag{3}$$

$$\hat{Y}_{21}^m = \{X_2^{m\circ}\hat{\beta}_1^m(\tau_1, z_1)e_k \quad \dots \quad X_2^{m\circ}\hat{\beta}_1^m(\tau_T, z_1)e_k\} \tag{4}$$

Kernel density estimates then show how the distributions vary depending on the values of the explanatory variables and the estimated coefficients.

The CPAR estimates provide additional information: how much of the difference in the predictions is due to differences in the locations of the observations across the two samples? For a given value of τ, the extended version of the decomposition that accounts for differences in the values of *z* is:

(Total) $$\hat{Y}_1^m(\tau, z_1) - \hat{Y}_2^m(\tau, z_2) = X_1^{m\circ}\hat{\beta}_1^m(\tau, z_1)e_k - X_2^{m\circ}\hat{\beta}_2^m(\tau, z_1)e_k = \tag{5a}$$

(Variables) $$X_1^{m\circ}\hat{\beta}_1^m(\tau, z_1)e_k - X_2^{m\circ}\hat{\beta}_1^m(\tau, z_1)e_k + \tag{5b}$$

(Location) $$X_2^{m\circ}\hat{\beta}_1^m(\tau, z_1)e_k - X_2^{m\circ}\hat{\beta}_1^m(\tau, z_2)e_k + \tag{5c}$$

(Coefficients) $$X_2^{m\circ}\hat{\beta}_1^m(\tau, z_2)e_k - X_2^{m\circ}\hat{\beta}_2^m(\tau, z_2)e_k \tag{5d}$$

This decomposition requires one additional set of estimated coefficients, $\hat{\beta}_1(\tau, z_2)$, which represents the estimated coefficients using the data from sample 1 evaluated at the sample 2 locations. All that is necessary to produce these estimates is to interpolate the estimated sample 1 coefficients to the locations represented in the second sample.[5] The order of the decomposition can be varied easily.

[5] An alternative to interpolation – using each location in sample 2 as a target point for estimation using the sample 1 data – may be preferable in situations where locations differ sufficiently to make interpolation unreliable.

4. Data and Model Specification

The full data set comprises nearly 235,000 sales of condominiums in Tokyo for 1986 – 2016. The dependent variable for our regressions is the natural log of the sale price per square meter of floor space. Explanatory variables include the log of floor area, the age of the unit, and a variable indicating that the unit has a southern view. We also include controls for the quarter of sale and the census tract.[6] Finally, we include the story on which the unit is located, along with variables indicating that the unit is on the first or second story. Preliminary data analysis suggested that the effect of the story on the sale price per square meter is close to being linear beyond the second floor. Our base estimating equation is a standard one for hedonic housing models: $y_i = x_i'\beta + d_i'\delta + s_i'\gamma + u_i$, where y represents the log of price per square meter; x includes the log of floor area, age, indicators of a southern view, and the story, along with the first and second story indicator variables; d is a group of variables indicating the quarter of sale; and s is the a group of variable indicating the census tract for the building. Sales prices are not adjusted for inflation, but the rate of inflation was low throughout this period.

Since the number of census tracts is very large (2,184 in 1986 – 1990 and 2,351 in 1991 – 1995), we follow a procedure proposed by Canay (2011) and first estimate the location effects, $\hat{\gamma}$, by a standard fixed effects regression. We then treat subtract $s_i'\hat{\gamma}$ from y_i before estimating any quantile regression. Thus, the dependent variable should be considered $y_i - s_i'\hat{\gamma}$ for the remainder of the paper.

Figure 1 presents the price index implied by 50% quantile regression estimates using the full data set. The price index is simply the estimated values of δ, with a base of zero for 2000:1.

[6] The data set is drawn from a region of 534 square kilometers. There are 2789 census tracts in the full data set, which implies an average area of 0.19 square kilometers, or 0.12 square miles. We also experimented with specifications using building (or "tower") fixed effects, but the tower indicator is not available before 1990 and approximately 15% of the towers had only one sale over the full time. Approximately half of the buildings for which the tower indicator is available had only one sale during the 1986-1995 period that is the focus of our analysis.

The variation of sales prices over time is striking. The value for the price index is -0.221 in 1986:1, or approximately 22% lower than the level in 2000:1. Prices rose dramatically over the next five years, with the index peaking at 0.911 in 1990:4. Prices then declined until the trough of -.037 was finally reached 2001:4. The rate of decline was particularly high from 1991 – 1995, with the index declining from 0.911 to 0.176 – a fall of more than 70% in 5 years. The value for the index is 0.402 for 2015:4, or approximately the same as its value in early 1987.

For the remainder of the paper, we focus on two particularly interesting periods – the striking rise in prices from 1986 – 1990 and the period of rapid decline from 1991 – 1995. Summary statistics for these two periods are presented in Table 1. Sales prices average 10.472 million yen for 1986 – 1990, compared with 8.808 million yen for 1991 – 1995. Average floor areas increased significantly across the two periods, rising from 50.466 square meters for 1986 – 1990 to 54.249 square meters in 1991 – 1995. Naturally, the average age also rose over time, from 9.721 years in 1986 – 1990 to 12.864 years for 1991 – 1995. Although new units were built during both periods, the rate of construction slowed considerably as prices declined: 24.22% of the units sold during 1986 – 1990 were no more than five years old, compared with 10.46% of the units sold in 1991 – 1995. Significantly more units sold had a southern exposure in 1991 – 1995. Stories range from 1 – 25, with no significance differences across periods.

5. Empirical Results

The base OLS results are shown in the first column of Table 1. The results indicate that the price per square meter is lower for larger units. Prices also decline with the age of the unit, and are lower on the first and second story of a building than on higher floors. Apart from this initial discount, prices rise on subsequent floors. A south view is associated with higher sales prices in the 1991 – 1995 period, but the estimate is statistically insignificant for 1986 – 1990. The regressions fit the data well, with R^2s of 0.899 for 1986 – 1990 and 0.833 in 1991 – 1995.

The remaining columns of Table 2 present quantile regression results for the 10%, 50%, and 90% quantiles, along with the difference between the 10% and 90% quantiles. Figure 2 shows how the coefficients for the log of floor area, age, a southern view, and the continuous story variable vary by quantile. The negative effect of floor area on the price square meter is higher in magnitude at higher quantiles, which suggests that the variability of per-meter sales prices is lower for larger units. In contrast, the negative effect of age is higher at low quantiles, which implies that the variability of per-meter sales prices is higher in older buildings. Our expectation would be that the coefficients for story would be higher at lower quantiles because the greater likelihood of a good view on higher stories should translate into lower variability of sales prices on higher floors. However, the coefficients for story do not have a clear pattern across quantiles.

Figure 3 displays the coefficients for the 10%, 50%, and 90% quantile regression. Since the 1986 – 1990 regressions use 1990:4 as the base and the 1991 – 1995 regression have 1991:1 as the base, these plots can be interpreted as 10%, 50%, and 90% quantile price indices, and the median results are simply a rescaled subset of the results shown in Figure 1. The results are remarkable for how close they appear to being parallel. However, Figure 4 shows that the spread between the 90% and 10% quantile widened during the first two years of the price boom: prices rose much more rapidly at the 90% quantile than at the 10% quantile, after which the rate of price growth for the 10% quantile began to catch up. Overall, these results are much different from the results found for the United States by Landvoigt, Piazzesi, and Schneider (2015) and McMillen (2016), who find that prices tended to rise more rapidly in low-priced regions of urban areas during the housing boom of the early 2000s. However, the pattern is consistent with the results of Deng, McMillen, and Sing (2012), who find that the distribution of residential prices shifted farther to the right for high-priced homes than for low-priced homes during times when prices in Singapore were rising rapidly. The results are also consistent with those found for Sydney, Australia by Waltl (forthcoming).

Previous studies such as Deng, McMillen, and Sing (2012); Landvoigt, and Schneider (2015); McMillen (2016), and Waltl (forthcoming) have found that appreciations can vary markedly across locations within an urban area. The overall shift in the distribution of the log of per meter sales prices across the two times is shown in Figure 5. The distribution shifted well to the left from 1986 – 1990 to 1991 – 1995, with a large increase in the area left of the central tendency. To allow for spatial variation in the shift in the distribution, we estimate CPAR versions of the quantile regression models. We use a tri-cube kernel with a 25% window size. The weights applied to each observation are a declining function of the straight-line distance between each observation and the grid cell midpoints. The coefficients are then interpolated to all locations in the data set. We estimate separate models for each period, for quantiles ranging from 0.04 to 0.96 in increments of 0.02. The estimates are stored in matrices with dimensions $n \times k \times T$, where n = 32,029 for 1986 – 1990 and n = 62,125 for 1991 – 1995, while k = 26 and T = 47 for both samples.

The spatial variation in the implied appreciation rates for 1986 – 1990 and 1991 – 1995 is shown in Figure 6, along with the difference across periods (1986 – 1990 *minus* 1991 – 1995). The estimated appreciation rates for 1986 – 1990 are the coefficients on the 1990:4 variable, as the base is 1986:1. Similarly, the estimated appreciation rates for 1991 – 1995 are the coefficients for the 1995:4 variable, with 1991:1 as the base. Figure 6 shows that there is substantial spatial variation in the estimated median appreciation rates. The estimated median appreciation rates for 1986 – 1990 range from 1.084 to 1.308 (i.e., 108.4% to 130.8%) while the estimated appreciation rates for 1991 – 1995 range from -0.943 to -0.564 (i.e., a decline ranging from 56.4% to 94.3%). The difference between the rates varies from 1.666 to 2.226. For the earlier period, the appreciation rates are highest for areas near downtown Tokyo and the Tokyo Harbor. These are the same areas where prices declined most rapidly in 1991 – 1995, and consequently these areas have the largest spreads between the growth rates across the two periods.

The patterns for the 10% quantile appreciation rates are similar to the median: appreciation rates are highest near downtown Tokyo in 1986 – 1991 with a range of 1.013 to 1.336, and the subsequent rates of declines are also highest in these areas, with a range of -0.966 to -0.549. The range for the absolute differences in 10% quantile appreciation rates over the two periods is 1.094 to 1.593 to 2.207. The patterns are somewhat different for the 90% quantile. While the area near downtown Tokyo is still the region with the highest rates of appreciation in 1986 – 1990, a large portion of this region had relatively low depreciation rates in 1991 – 1995. However, the rates of appreciation were so high in the earlier period that the absolute differences between the rates of change are again highest in this region: although the deprecation rates were somewhat modest in 1991 – 1995, they were paired with extremely high rates of appreciation in 1986 – 1990. Across the full region, the range of 90% quantile appreciation rates is 1.094 to 1.318 in 1986 – 1990, -0.966 to -0.525 in 1991 – 1995, and the range of absolute differences is 1.629 – 2.257.

Although the CPAR quantile approach generates a seemingly overwhelming number of coefficient estimates, the results are easy to interpret by calculating the implied marginal distribution of the dependent variable at suitable values of the explanatory variable. We focus on the four primary explanatory variables – the log of floor area, age, a southern view, and the unit's story.[7] We set values for the log of floor area to 3.4, 3.9, and 4.2, which correspond (after rounding) to 30, 50, and 70 square meters. The other variables are set to their actual values in the data for these calculations. We then make comparable calculations for other variables: age is set to 10, 20, and 30 years; a southern view is set to 0 and 1, and the story is set to 1, 10, and 20. When the story is set to 1, the value for the 1st-story variable is also set to 1.

The results for the four primary explanatory variables are shown in Figure 9. Increased floor area shifts the distribution of log per meter sale price to the left, with lower variability at higher areas. Increasing age from 10 to 20 to 30 years shifts the distribution markedly to the left, with little discernible effect on the spread of the distribution. A southern view has little effect on the distribution of sales prices. Moving from the 1st story to 10 and 20 shifts the distribution of prices well to the right, but again with little effect on the variance of the distribution.

Figure 10 displays the results of a series of similar calculations for the quarter of sale. The other explanatory variables are set to their actual values in data set, but when the variable for one quarter is set to 1, the values for all other quarters are set to 0. Figure 10 shows the results for the first quarters of 1986, 1988, 1991, 1993, and 1995. The distribution of the log of per meter sale price shifts far to the right from 1986 to 1988 to 1991, with a marked increase in the variance. From 1991 to 1993, the center of the distribution shifts all the way back to the 1988 point, with a reduction in variance that roughly matches the original 1986 level. The distribution

[7] For these counterfactual calculations, we pool the 1986 – 1990 and 1991 – 1995 data sets, which imposes a restriction that the coefficients for these variables are constant for a given quantile over time. The advantage of pooling the data is to isolate the effects of changes in the value of the explanatory variable rather than combining the effects of changes in the variable and changes in the coefficients. The models are again estimated using a 25% window and a tri-cube kernel for the straight-line distance between each observation and the target points.

shifts further still to the left in 1995, with an increase in the variance. Between 1993 and 1995, the upper end of the distribution clearly shifts further than the lower end.

6. Decompositions

Figure 5 shows that the distribution of sales prices shifted far to the left between 1986 – 1990 and 1991 – 1995 in Tokyo. How much of this change was due to changes in the explanatory variables, changes in the location of the sales, and changes in the coefficients? Before presenting our decomposition results, note that the descriptive statistics from Table 1 and the graphs of the quantile regression estimates in Figure 2 provide some insight into the patterns that might be expected. Since all existing buildings are older in the later period, the change in the age variable will tend to shift the distribution of sales prices to the left. However, Figure 2 suggests that this tendency will be at least somewhat ameliorated by the increase in the magnitude of the age discount for 1991 – 1995 relative to the earlier period. Second, Table 1 shows that the size of the units increases significantly over time, and this effect is reinforced by the tendency toward higher coefficient on the log of floor area in 1991 – 1995.

Our estimated models differ somewhat from our presentation in Section 3 in that many of the variables represent the quarter of sale. Since it is not possible to analyze how, for example, the values of the variable for 1987:1 are different in the later sample, a decomposition of the change in variable is irrelevant for these variables. If we redefine $d_i'\delta(z)$ as simply $D(z_i)$, the following two terms are added to the decomposition in equation (5):

(Location, Time of Sale) $$\widehat{D}_1^m(\tau, z_1)e_k - \widehat{D}_1^m(\tau, z_2)e_k + \quad (6e)$$

(Time of Sale) $$\widehat{D}_1^m(\tau, z_2)e_k - \widehat{D}_2^m(\tau, z_2)e_k \quad (6f)$$

Equation (6e) shows the effect of changing the location of sales from the period 1 to period 2 locations on the period 1 time coefficients. For examples, if the locations of the sales in period 2 are concentrated in areas that happened to have relatively lows rates of appreciation in the first period, the distribution of $\widehat{D}_1^m(\tau, z_2)e_k$ will be further to the left than the distribution implied by $\widehat{D}_1^m(\tau, z_1)e_k$. By holding the sales locations constant, equation (6f) then shows the pure effect of the time coefficients.

Figure 11 presents the components of the decomposition for the base explanatory variables, the log of floor area, age, a southern view, and the three story variables. The intercept is not included in this set of variables. The effect of changing the values of these variables to their 1991 – 1995 values is the change from "x1b1" to "x2b1" in Figure 11. The age variable is the dominant one in this change: since buildings age over time, the direct effect of changing this variable to its 1991 – 1995 values is to shift the distribution of sales prices to the left. Next, the effect of changing the locations of the sales from the 1986 – 1990 sites to those observed for 1991 – 1995 is the shift from "x2b1" to "x2b12" in Figure 11. The effect of this change is to move the distribution of prices still further to the left, except in the very low-priced portion of the distribution. Finally, the change in the estimated coefficients to their 1991 – 1995 values shifts the distribution back to the right – close to the 1986 – 1995 starting point, but somewhat further to the left. Thus, the change in coefficients offsets most of the change in the variables and the location of the sales.

Figure 12 adds the effect of the quarter of sale variables to these sale price densities. The base is the combination of predicted values from the other explanatory variables for 1991 – 1995 (“x2b2”) and the period 1 quarter of sale predictions at their period 1 locations (“D11”).[8] Evaluating these quarter of sale values at the 1991 – 1995 sales locations (“x2b2 + D12) has very little effect on the distribution of sales prices. Switching the coefficients to their 1991 – 1995 values (“x2b2 + D22”) shifts the distribution of sales prices well to the right, although the fat right tail disappears in the process.

[8] The intercept again is not included in these calculations. Since the base period for the second period is 1991:1, adding the intercept shifts the distributions in Figure 12 so far to the right that the difference between the other functions is obscured.

7. Conclusion

Conditionally parametric estimators are a convenient away to allow for spatial variation in model coefficients when a global parametric specification is not suitable for the entire area covered in a data set. The models are estimated by placing more weight on observations close to a set of target points. After either estimating the model for every location represented in a data set or interpolating from the target points to other locations, CPAR models produce a separate set of coefficients for every point in the data set. Although the large number of coefficients would appear to make the results difficult to interpret, counterfactual distributions can easily be used to display the results for discrete changes in the values of an explanatory variables, which readily summarize the direction of and magnitude of a variable's effect on the overall distribution of the dependent variables.

In this paper, we extend Machado and Mata's (2005) method for decomposing changes in the distribution of the dependent into the portions explained by differences in the explanatory variables and coefficients by allowing also for changes in the location of the observations represented in different samples. The procedure requires only one additional sets of calculations – using the data from one sample to estimate the model at the locations represented in the other sample. The procedure can be applied to either standard CPAR (or "geographically weighted regression") estimates or to CPAR quantile models.

We use the CPAR quantile approach to analyze changes in the distribution of per meter sales prices for condominiums in Tokyo for 1986 – 1990 and 1991 – 1995, which represent periods when prices first rose dramatically, followed by a sharp downturn. The decompositions suggest a large portion of the change in change in distributions across the two periods is due to changes in the variables. The variable with the most influence on the change in distributions is the age of the unit, which naturally tends to become larger over time. Although higher ages shift the distribution of sales prices to the left, changes in the floor area have the opposite effect because newly constructed building tended to be larger than the existing units in Tokyo at the time. Changes in the coefficients offset much of this leftward shift in the price distribution as the discount associated with older units declined from 1986 – 1990 to 1991 – 1995.

The CPAR estimates reveal significant spatial variation in the appreciation rates of sales prices from 1986 – 1990, and then again for the large rates of depreciation observed for 1991 – 1995. Prices rose most rapidly near downtown Tokyo, and subsequently declined most rapidly in the same areas. However, the patterns were not uniform across time periods: the 90% quantile price indices reveal much lower depreciation rates near downtown Tokyo for 1991 – 1995 than is implied by the 10% or 50% quantile. Taking the quarter of sale into account in the price distribution decompositions suggests that sales for 1991 – 1995 were located in areas that had relatively low rates of depreciation in 1986 – 1990, which suggests that standard decomposition would have attributed some of the effects of location to changes in the coefficients or variables.

References

Canay, Ivan A. 2011. "A Simple Approach to Quantile Regression for Panel Data," *Econometrics Journal*, 14, 368-386.

Carrillo, Paul and Anthony Yezer. 2009. "Alternative Measures of Homeownership Gaps across Segregated Neighborhoods," *Regional Science and Urban Economics*, 39, 542-552.

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

Chernozhukov, Victor and Christian Hansen. 2006. "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491-525.4

Cleveland, W. S. 1994. "Coplots, Nonparametric Regression, and Conditionally Parametric Fits." In T.W. Anderson, K.T. Fant, and I. Olkin (Eds.), *Multivariate Analysis and its Applications*, pp. 21-36. Hayward: Institute of Mathematical Statistics.

Cleveland, W. S., E. H. Grosse, and W. M. Shyu. 1992. "Local Regression Models. In J. M. Chambers and T. J. Hastie (Eds.), *Statistical Models in S*, pp. 309-376. Pacific Grove: Wadsworth and Brooks/Cole.

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

Kim, Tae-Hwan and Christophe Muller. 2004. "Two-Stage Quantile Regression when the First Stage is Based on Quantile Regression," *Econometrics Journal*, 7, 218-231.

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

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

Landvoigt, T., M. Piazzesi, and M. Schneider. 2015. "The Housing Market(s) of San Diego," *American Economic Review*, 105, 1371-1407.

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

Loader, Clive. 1999. *Local Regression and Likelihood*. New York: Springer.

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

McMillen, Daniel P. 2013. *Quantile Regression for Spatial Data*, Springer Briefs in Regional Science, New York.

McMillen, Daniel. 2015. “Conditionally Parametric Quantile Regression for Spatial Data: An Analysis of Land Values in Early Nineteenth Century Chicago,” *Regional Science and Urban Economics*, 55, 28-38.

McMillen, Daniel. 2016. “Local Quantile House Prices,” manuscript, University of Illinois at Urbana-Champaign.

Oaxaca, Ronald. 1973. “Male-female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 14, 693-709.

Waltl, Sofie R. Forthcoming. “Variation Across Price Segments and Locations: A Comprehensive Quantile Regression Analysis of the Sydney Housing market,” *Real Estate Economics*.

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

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.

Zhang, Lei. 2016. “Flood Hazards’ Impact on Neighborhood House Prices: A Spatial Quantile Regression Analysis,” *Regional Science and Urban Economics*, 60, 12-19.

Zhang, Lei and Tammy Leonard. 2014. “Neighborhood Impact of Foreclosure: A Quantile Regression Approach,” *Regional Science and Urban Economics*, 48, 133-143.

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
1986 – 1990 (32,029 obs.)				
Price (10,000 yen) per Square Meter	104.721	47.082	19.188	339.971
Log Price per Square Meter	4.553	0.450	2.954	5.829
Floor Area (square meters)	50.466	17.332	16.000	133.900
Log Building Area	3.854	0.384	2.773	4.897
Age	9.721	5.439	0.416	31.916
South View	0.209	0.407	0	1
1^{st} Story	0.098	0.297	0	1
2^{nd} Story	0.158	0.364	0	1
Story	4.753	2.951	1	25
1991 – 1995 (62,125 obs.)				
Price 10,000 yen) per Square Meter	88.079	36.308	19.569	348.052
Log Price per Square Meter	4.408	0.363	2.974	5.852
Floor Area (square meters)	54.249	18.609	16.000	134.550
Log Building Area	3.928	0.379	2.773	4.902
Age	12.864	5.968	0.416	34.416
South View	0.383	0.486	0	1
1^{st} Story	0.097	0.295	0	1
2^{nd} Story	0.157	0.364	0	1
Story	4.766	3.013	1	25

Table 2: OLS and Quantile Estimates, 1986 – 1990

Variable	OLS	Quantile 0.1	Quantile 0.5	Quantile 0.9	Difference, Quantile .9 - .1
1986 – 1990 (32,029 obs.)					
Log Floor Area	**-0.0431** (0.0028)	**-0.0258** (0.0038)	**-0.0519** (0.0022)	**-0.0584** (0.0041)	**-0.0326** (0.0040)
Age	**-0.0223** (0.0002)	**-0.0246** (0.0003)	**-0.0223** (0.0002)	**-0.0218** (0.0003)	**0.0028** (0.0004)
South View	0.0048 (0.0025)	0.0056 (0.0040)	**0.0088** (0.0023)	-0.0009 (0.0044)	-0.0064 (0.0037)
1st Story	**-0.0140** (0.0035)	**-0.0174** (0.0057)	**-0.0168** (0.0033)	-0.0116 (0.0063)	0.0058 (0.0085)
2nd Story	**-0.0061** (0.0028)	-0.0040 (0.0046)	**-0.0062** (0.0027)	-0.0079 (0.0051)	-0.0039 (0.0070)
Story	**0.0062** (0.0004)	**0.0085** (0.0006)	**0.0061** (0.0004)	**0.0044** (0.0007)	**-0.0041** (0.0008)
1991 – 1995 (62,125 obs.)					
Log Floor Area	**-0.0440 (0.0020)**	**0.0104 (0.0026)**	**-0.0472 (0.0017)**	**-0.0686 (0.0030)**	**-0.0790 (0.0048)**
Age	**-0.0221 (0.0001)**	**-0.0233 (0.0002)**	**-0.0221 (0.0001)**	**-0.0213 (0.0002)**	**0.0020 (0.0002)**
South View	**0.0135 (0.0014)**	**0.0132 (0.0020)**	**0.0158 (0.0013)**	**0.0104 (0.0023)**	**-0.0028 (0.0026)**
1st Story	**-0.0117 (0.0025)**	**-0.0128 (0.0038)**	**-0.0129 (0.0025)**	**-0.0109 (0.0045)**	**0.0020 (0.0052)**
2nd Story	**-0.0063 (0.0020)**	**-0.0043 (0.0031)**	**-0.0072 (0.0020)**	**-0.0082 (0.0036)**	**-0.0039 (0.0040)**
Story	**0.0057 (0.0003)**	**0.0065 (0.0004)**	**0.0055 (0.0003)**	**0.0049 (0.0005)**	**-0.0016 (0.0006)**

Notes. Significance at the 5% level is indicated by bold face. The regressions also include 20 variables indicating the quarter of sale and census area fixed effects (2,184 in 1986 – 1990 and 2,351 in 1991 – 1995). The R^2s for the OLS models are 0.899 in 1986 – 1990 and 0.833 in 2991 – 1995.

Figure 1: Quantile Median Price Index, 1986 – 2015

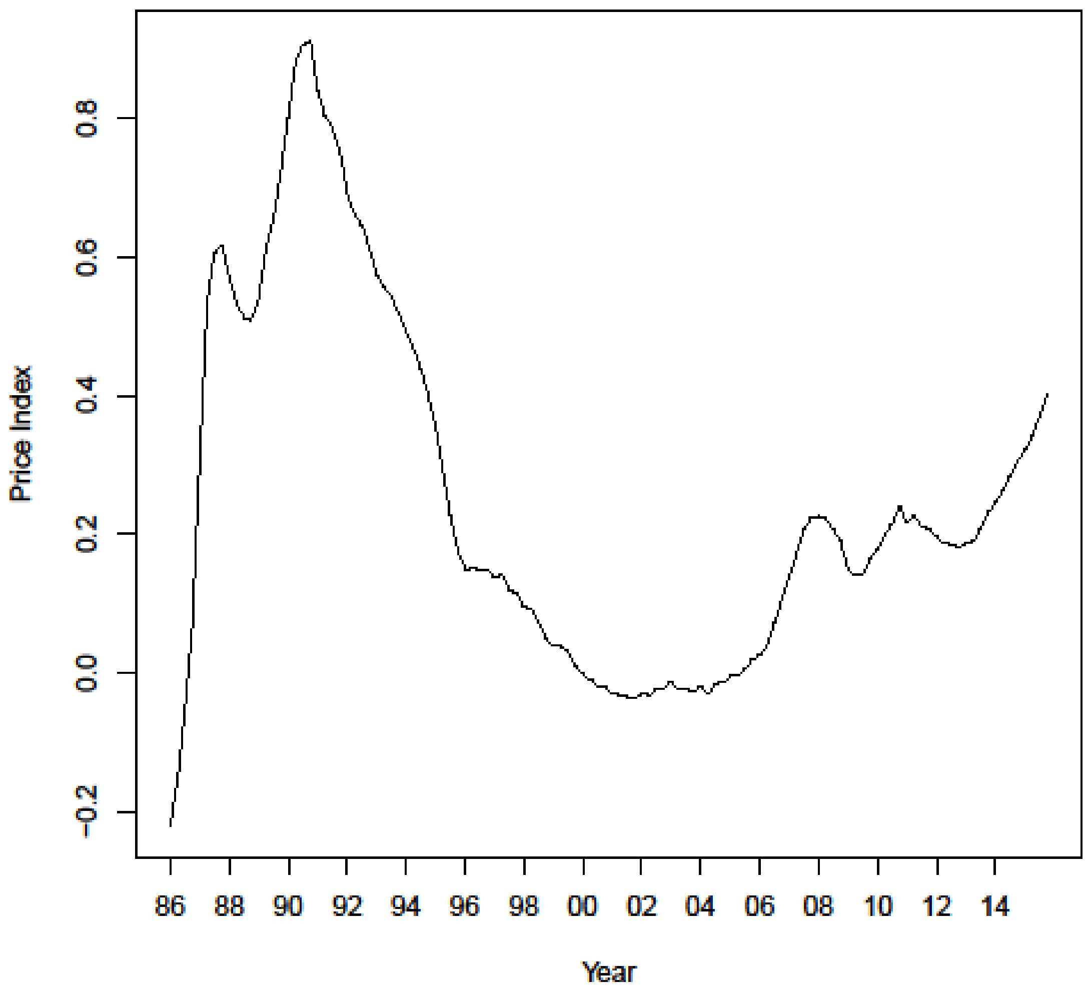


Figure 2: Coefficient Estimates by Quantile

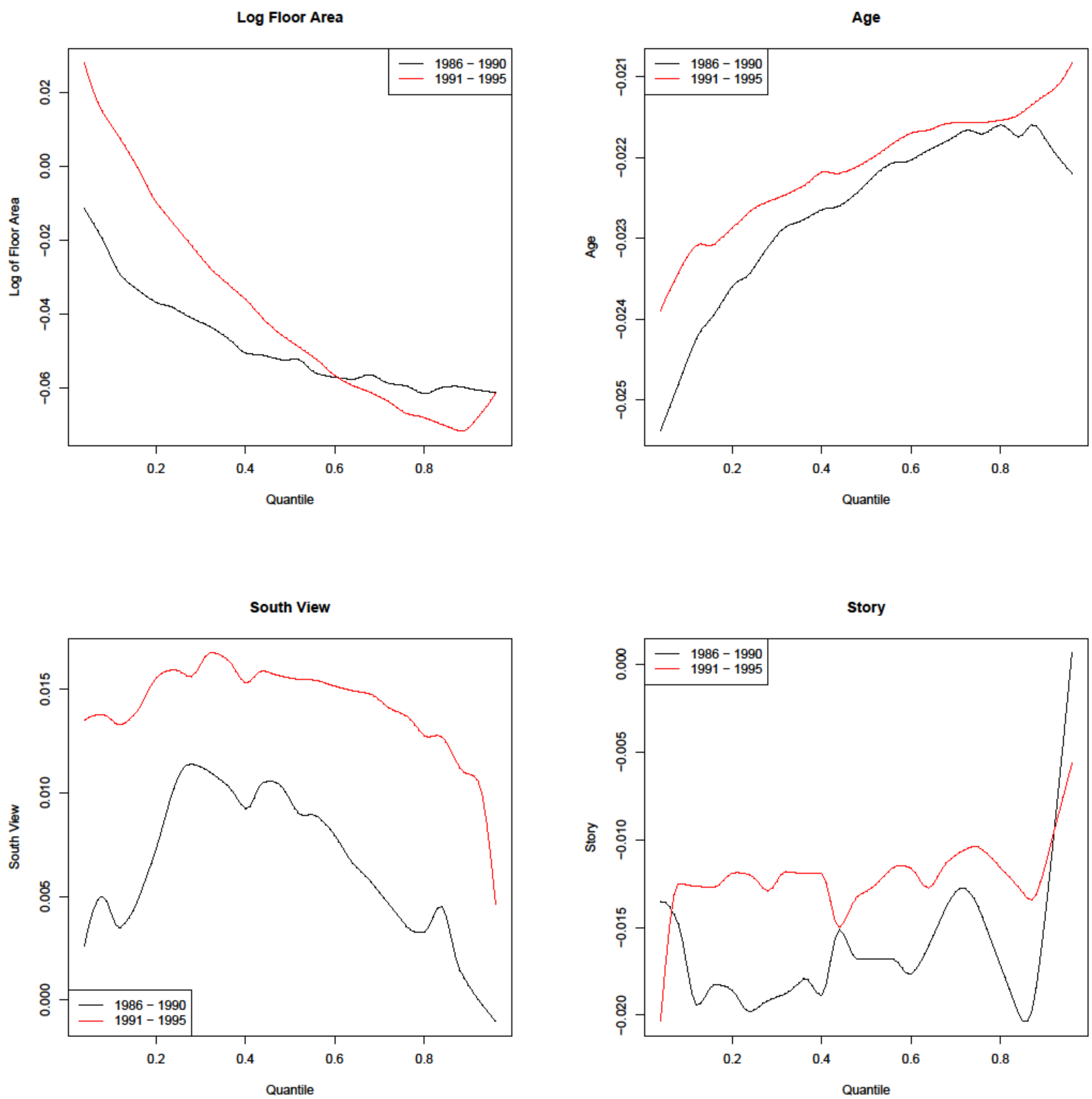


Figure 3: 10%, 50%, and 90% Quantile Estimates, 1986 – 1995

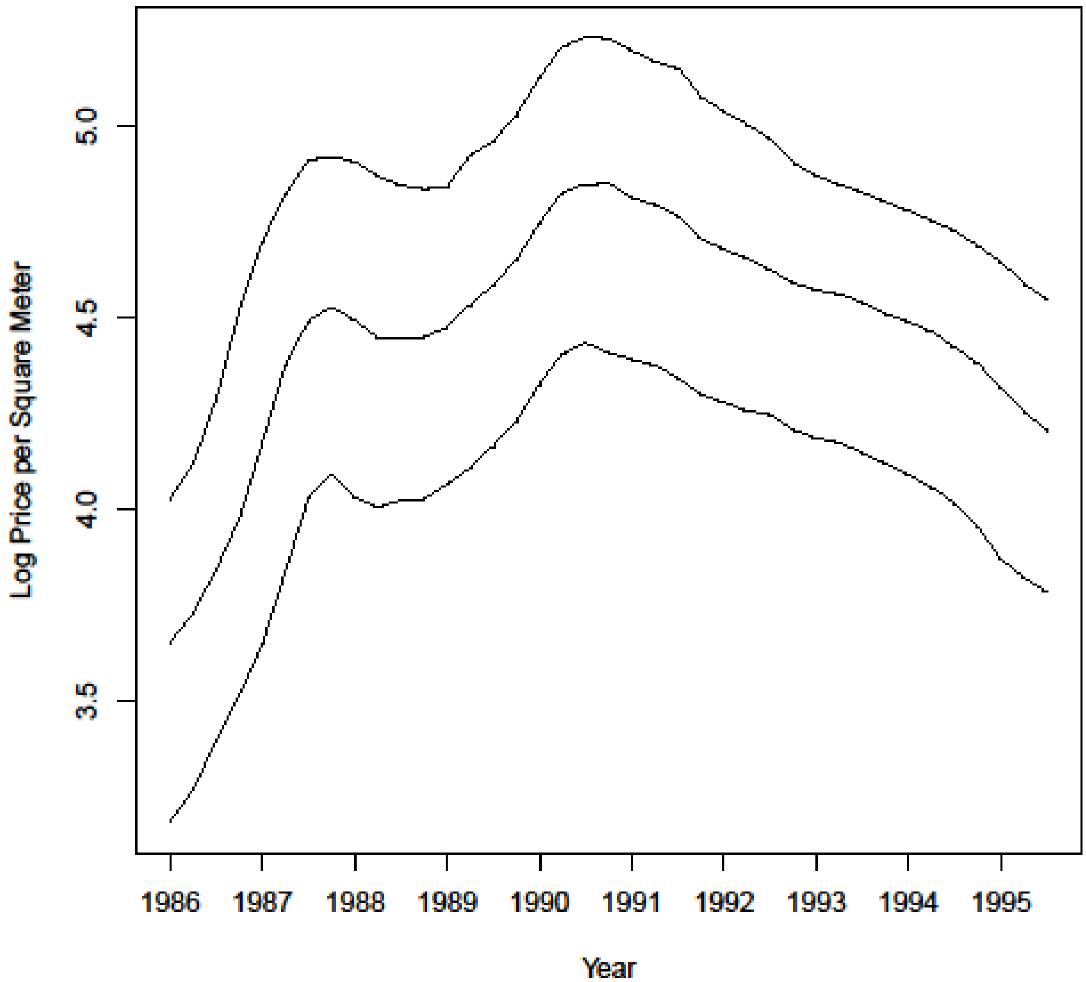


Figure 4: Difference between 90% and 10% Quantile Estimates, 1986 – 1995

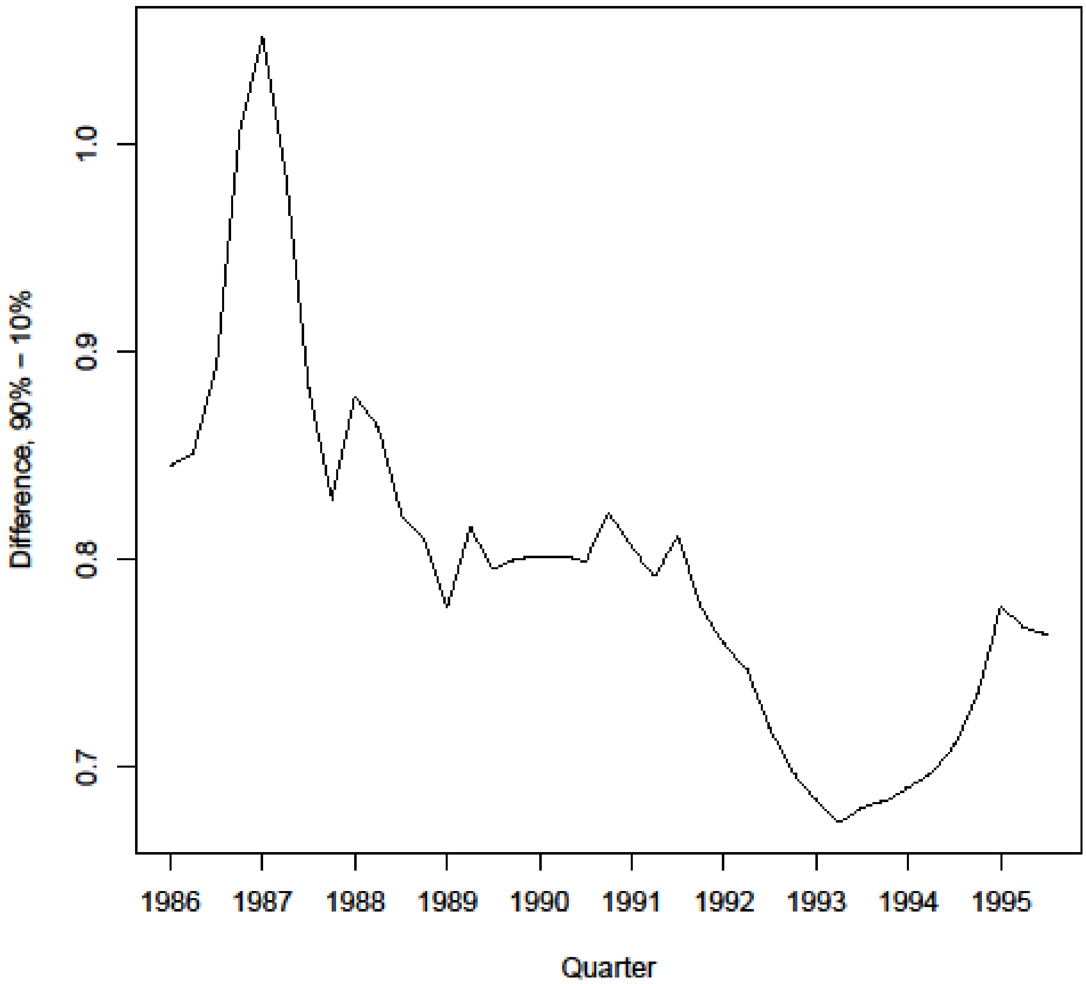


Figure 5: Kernel Density Estimates for Log Price per Square Meter

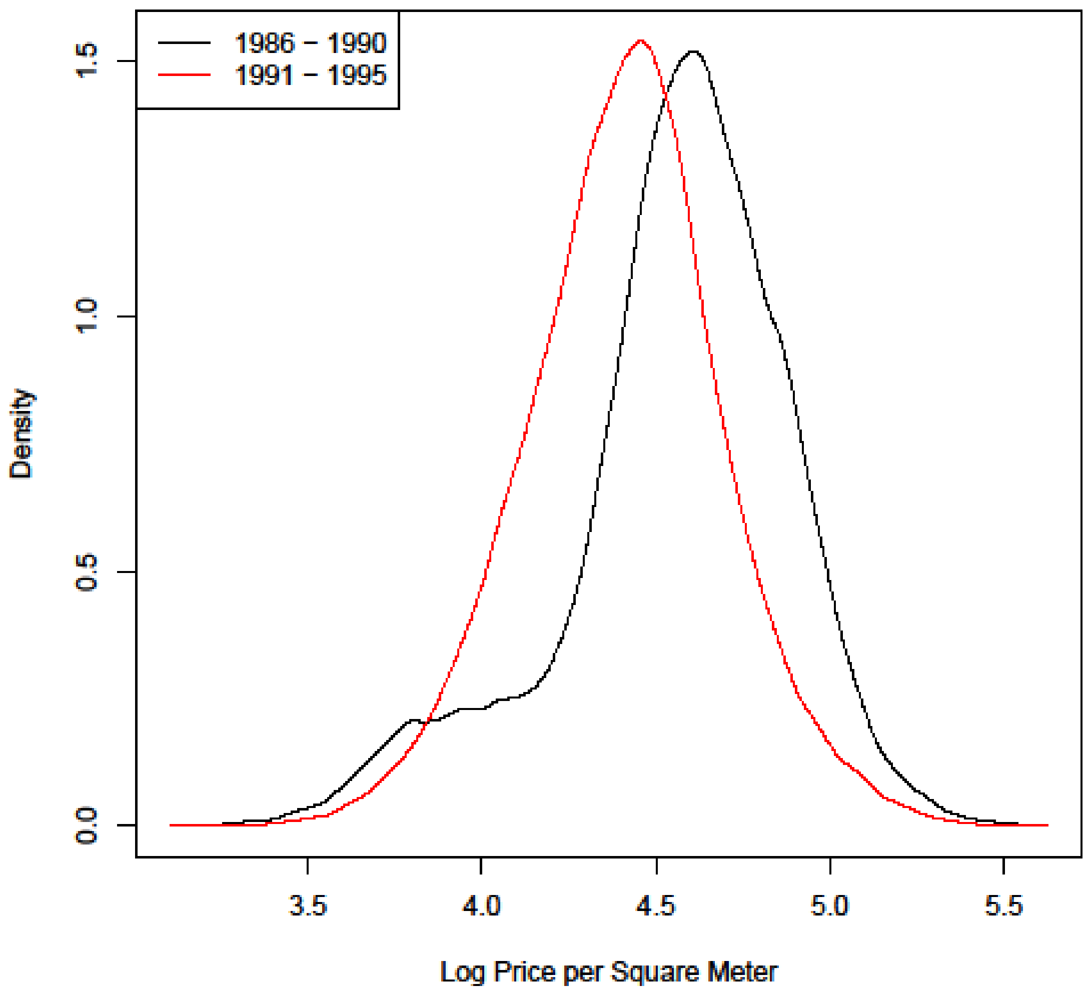
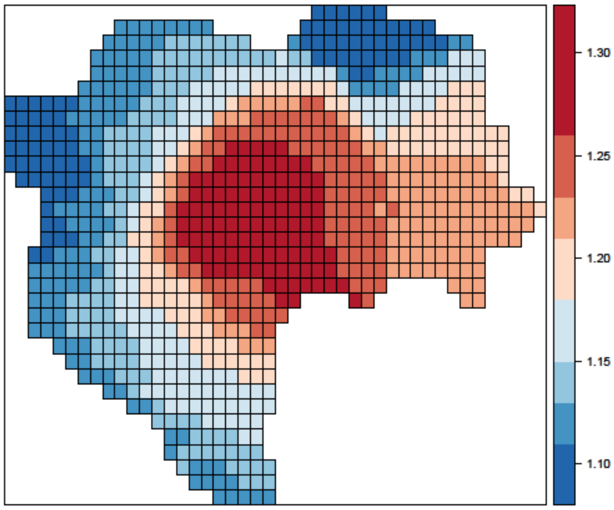
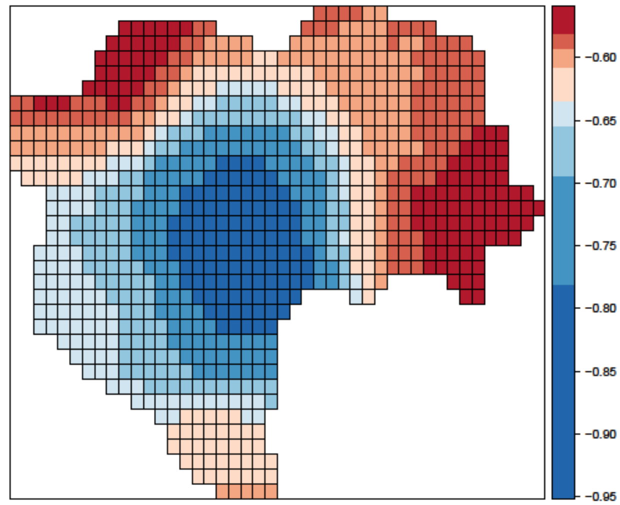


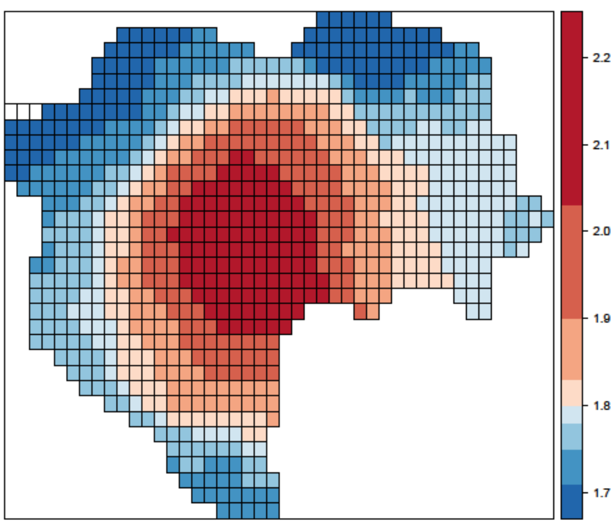
Figure 6: Spatially Varying Median Appreciation Rates

1986 – 1990





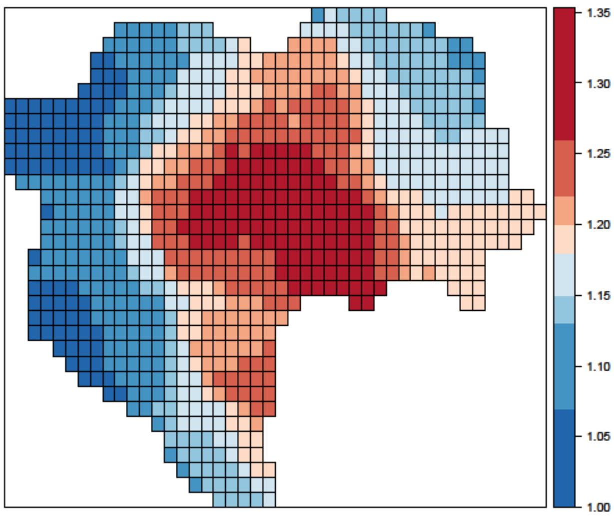
Difference, 1986-90 – 1991-2005

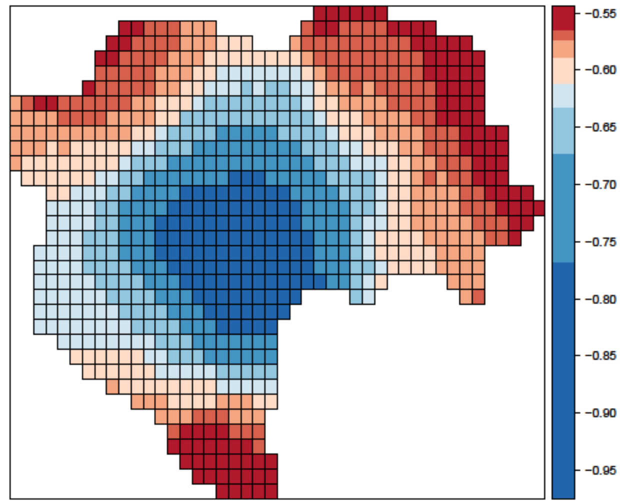


1991-1995

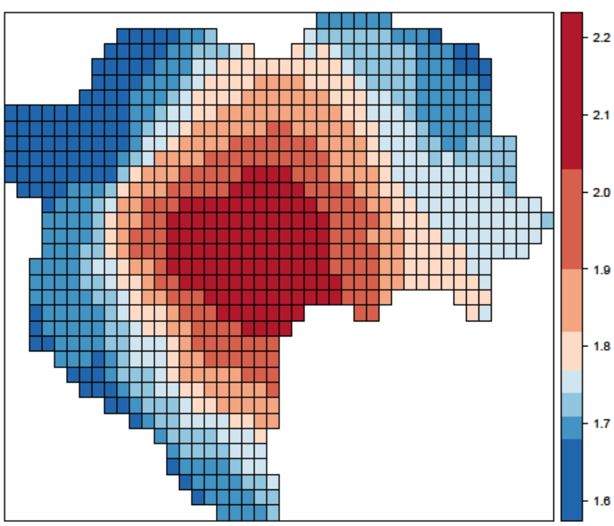
Figure 7: Spatially Varying 10% Quantile Appreciation Rates

1986 – 1990





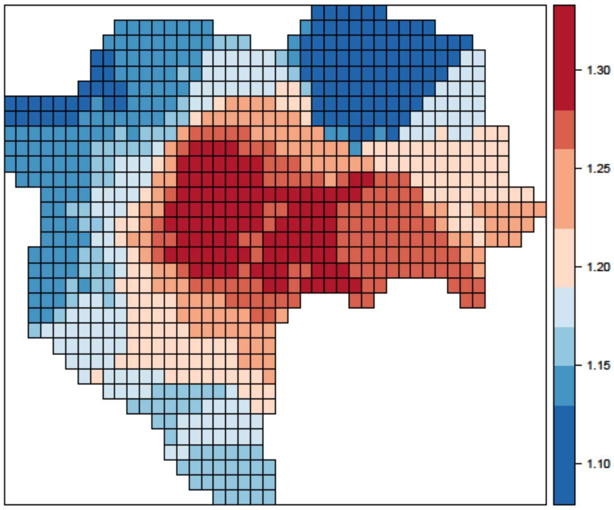
Difference, 1986-90 – 1991-2005



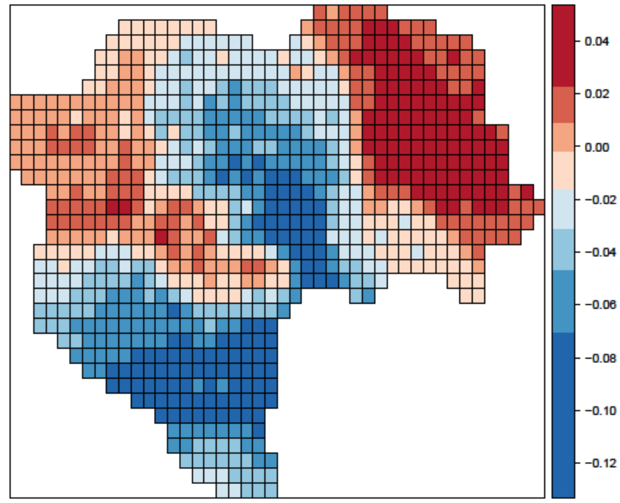
1991-1995

Figure 8: Spatially Varying 90% Quantile Appreciation Rates

1986 – 1990



1991-1995



Difference, 1986-90 – 1991-2005

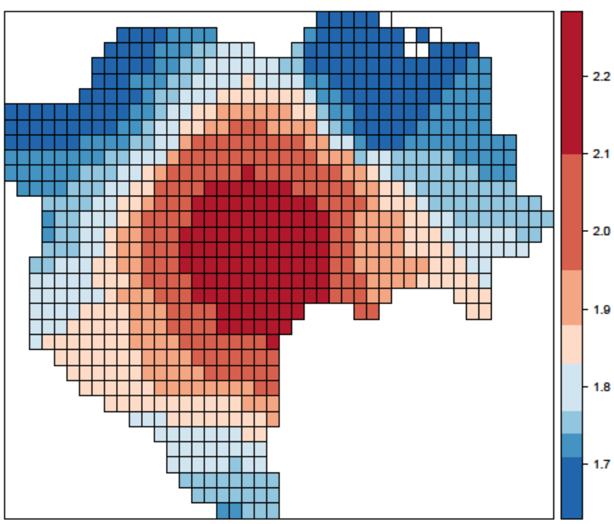


Figure 9: Predicted Effects of Discrete Changes in Variables

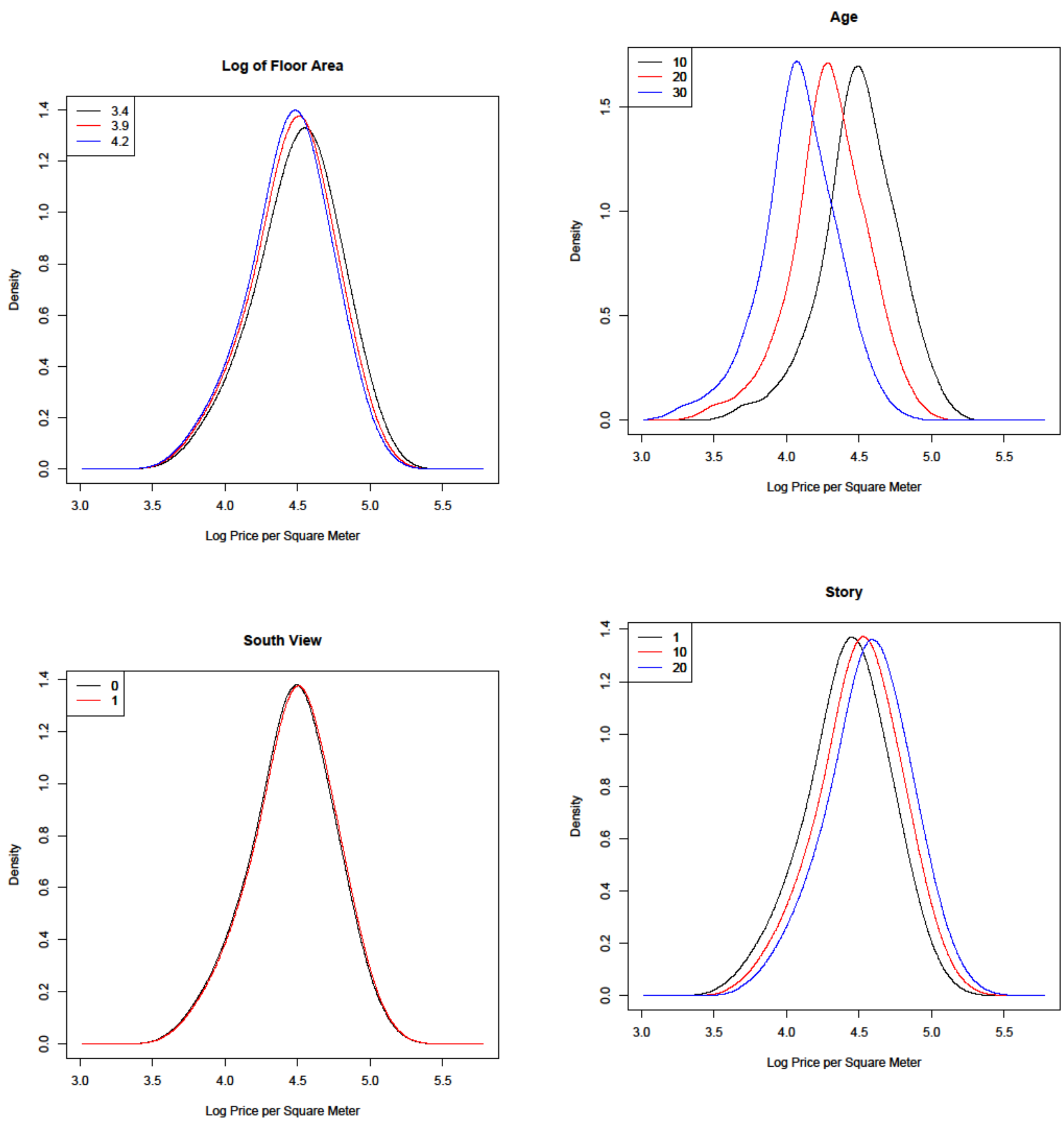


Figure 10: Predicted Effects Changes in Quarter of Sale

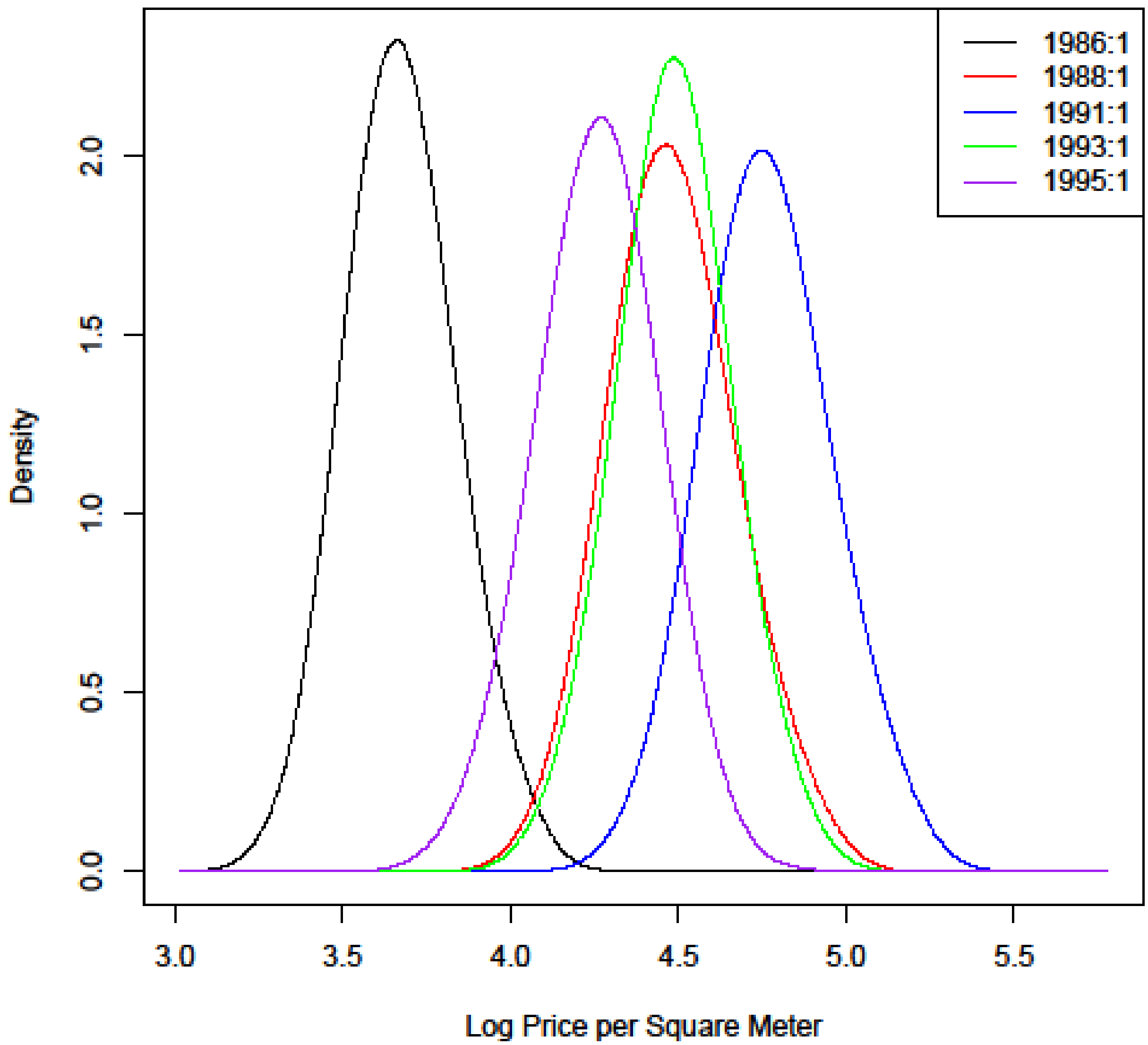


Figure 11: Decomposition 1986 – 1990 to 1991 – 1995, X

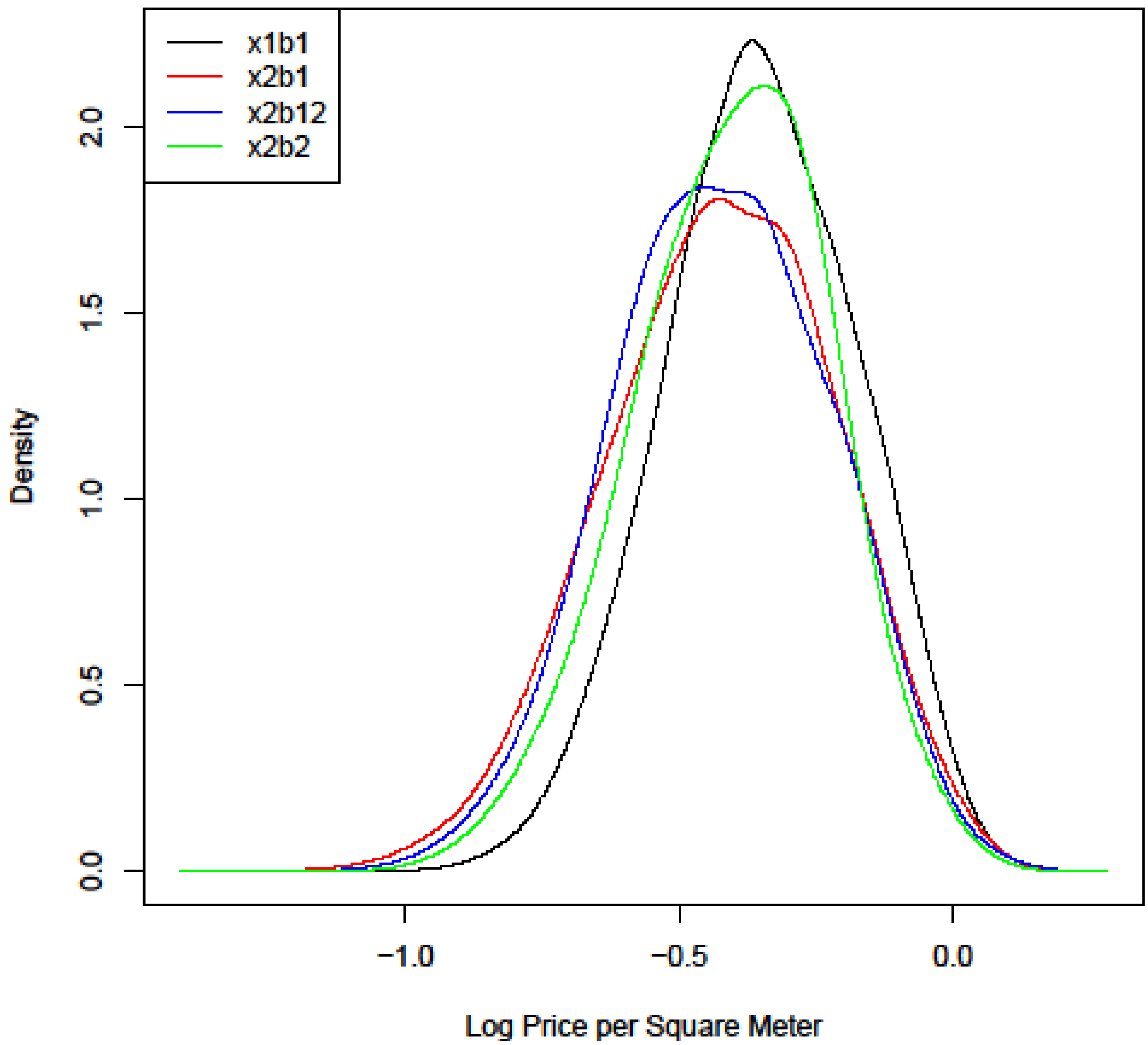


Figure 12: Decomposition 1986 – 1990 to 1991 – 1995, Time of Sale