Locally Weighted Quantile House Price Indices and Distribution in Japanese Cities, 1986–2015

Yongheng Deng[1], Xiangyu Guo[2], Daniel McMillen[3] and Chihiro Shimizu[4]

This Draft: March 27, 2017

Abstract

Locally weighted regression, or *loess*, is a non-parametric approach taken through a multivariate smoothing procedure; it allows the coefficients of hedonic house price regressions to vary over space. Using data on residential condominium sales in 15 Japanese cities over the 1986–2015 period, we construct house price indices based on *loess* and quantile regressions. The estimates show how the appreciation rate of house prices changed over time in the Greater Tokyo and Kansai areas, especially during the boom and the burst. One advantage of taking the locally weighted quantile approach is that it facilitates comparisons of the change in a full distribution for various large metropolitan areas, each of which contains multiple cities. During the Japanese asset bubble of the 1980s, house prices in the Greater Tokyo area rose more rapidly in the boom (and declined more slowly in the burst) than those in the Kansai area. The house prices in the Kansai area had a high appreciation rate during the boom and declined more afterwards, compared to the Greater Tokyo area. The distribution of house price changes was larger in the Kansai area, which saw a small degree of variation before the boom but had a large one at the peak of the boom. During the boom, the appreciation rate of high-priced houses was larger than that of low-priced houses in the Kansai area; meanwhile, in the Greater Tokyo area, the prices of high-priced houses appreciated less than did those of low-priced houses.

Key Words: house price indices; quantile regressions; locally weighted regressions; price distribution; asset bubble

JEL Classification: R21; R31; C10

---

[1] Departments of Finance and Real Estate, National University of Singapore, #04-02 I-Cube Building, 21 Heng Mui Keng Terrace, Singapore, 119613, Singapore (email: ydeng@nus.edu.sg)
[2] Department of Real Estate, National University of Singapore, #04-02 I-Cube Building, 21 Heng Mui Keng Terrace, Singapore, 119613, Singapore (email: xguo@u.nus.edu)
[3] Department of Economics, University of Illinois, 1407 W. Gregory St. Urbana, IL 61801, USA (email: mcmillen@illinois.edu)
[4] Nihon University, Setagaya Campus, 3-34-1 Shimouma, Setagaya-ku, Tokyo 154-8513 Japan, and Institute of Real Estate Studies, National University of Singapore, 21 Heng Mui Keng Terrace, #04-02, 119613 Singapore (email: shimizu.chihiro@nihon-u.ac.jp)

1. Introduction

In many economically advanced nations, the formation and collapse of property bubbles has had a profound impact on economic administration. In Japan, the property bubble that began around the mid-1980s has been called the greatest bubble of the 20th century. In the aftermath of the bubble's collapse, the country faced a period of long-term economic stagnation, which has been dubbed the "lost decade." Many other countries have had similar experiences concerning this type of problem; examples include Sweden's economic crisis in the 1990s, and the global financial crisis and economic stagnation caused by the formation and collapse of the US-centered property bubble in the early 21st century.

It is important to understand how house prices develop during a boom and a burst. House prices may not change at a uniform rate within a suburban area or across metropolitan areas. For example, the timing of the boom in the Greater Tokyo area may have varied from that of the Kansai area. Moreover, as each metropolitan area is large, the appreciation rates of high and low-priced houses might differ within a region; even the change in distribution can vary within a metropolitan area. In such cases, a single price index cannot adequately describe changes in house price distribution across metropolitan areas.

Mean-based price indices—such as hedonic and repeat sales—are based on the assumption that appreciation rates do not vary across urban areas or house-price levels. It is note inadequate merely to compare changes in median indices over different metropolitan areas. Under such circumstances, quantile regressions focus on the change in distribution; this allows for appreciation rates that are different for high and low-priced houses. Index values based on quantile regressions can easily compare changes in distribution across metropolitan areas. Both hedonic regression and quantiles regression assume that the coefficients of covariates are constant within suburban areas; however, as many metropolitan areas are large, it is more convincing to allow for the coefficient of covariates to vary over space. Locally weighted regression (loess) is a non-parametric approach that allows coefficients to vary smoothly for a neighborhood. Indices based on the locally weighted approach could describe more appropriately how property markets develop within small geographic areas.

The objective of this study is to use locally weighted and the quantile approach to construct house price indices, to show how distribution changed in both the Greater Tokyo area and the Kansai area in the 1986–2015 period. We then compare the difference in the change in distribution of these two metropolitan areas affected by the asset bubbles. Additionally, we could assign numbers typical to the characteristics of an apartment—such as floor space or building age—to illustrate the estimated counterfactual distribution.

The results of this study provide some insight into the rise and fall of house prices over boom and burst periods. During the Japanese asset bubble of the 1980s, house prices in the Greater Tokyo area rose earlier in the boom, and declined later in the burst, than was the case in the Kansai area. In comparison to those in the Greater Tokyo area, the house prices in the Kansai area had a large appreciation rate during the boom, and declined more afterwards. The distribution of house prices changed more markedly in the Kansai area, which saw little variation before the boom but saw large variation at the peak of the boom. During the boom itself, in the Kansai area, the appreciation rates of high-priced houses were larger than those of low-priced houses, while in the Greater Tokyo area, the prices of high-priced houses appreciated less than the low-priced houses did. During the burst, the price of low-priced houses declined earlier than did those of median and high-priced houses. The prices of houses in the Greater Tokyo area stopped declining in 2000, while the prices of houses in the Osaka area continued to decline until 2005.

2. Background

The most commonly used price indices are the hedonic price index and the repeat sales index. In the hedonic model, the log of the sale price is expressed as a function of certain characteristics of the structure, location and sale date. Structure characteristics include floor space, building age etc, while the covariates of the location contain include the distance to the subway, schools and parks etc. Time dummy variables are also included in the hedonic model, the coefficients of which reflect the price changes. The standard hedonic model is:

$$lnP_{it} = X_{it}\beta + \sum_{t=1}^{T} D_{it}\delta_t + u_{it} \quad (1)$$

where P represents the sale price, X is a combination of variables that represents structure and location characteristics, and D is the matrix of time dummies. The set of coefficients of D is the hedonic price index. Rosen (1974) points out that the correct functional form cannot be determined on theoretical grounds, unless there is a costless repackaging of the characteristics; thus, the hedonic model cannot sidestep the "omitted variable" problem. In addition, the hedonic model assumes that the coefficient of each variable is the same at different times and for different houses; however, the coefficients of the explanatory variable may change over time and over space.

Invoking an additional assumption—namely, that X_{it} is constant over time for each property—Bailey et al. (BMN 1963) propose the repeat sales price index. Subsequently, Case and Shiller's (1989) modification of that model became the most popular housing index of the US market.[5] The function formation is:

$$lnP_{it} - lnP_{is} = \delta_t - \delta_s + u_{it} - u_{is} \tag{2}$$

where $s < t$. The repeat sales approach requires that the same houses sell at least twice during the sample period. With hedonic estimation, the repeat sales estimator is often presented as a potential solution to the omitted bias problem, since structural and neighborhood features are constant over time. However, the sample of repeat sales

[5] It should be noted that an adjustment to cope with the age effect is made in constructing the official S&P/Case–Shiller home price index. Standard & Poor's (2008) states that "Sales pairs are also weighted based on the time interval between the first and second sales. If a sales pair interval is longer, then it is more likely that a house may have experienced physical changes. Sales pairs with longer intervals are, therefore, given less weight than sales pairs with shorter intervals" (p. 7).

houses may not be representative of all transactions within the property market. Moreover, the assumption that attributes are constant might be violated, due to the age problem and renovation issues. Case and Quigley (1991) construct a hybrid model to resolve the age problem.[6]

The repeat sales estimators are restricted for small geographic areas where the number of repeat-sales pairs is likely to be small. Wang and Zorn (1997) show that the repeat sales approach is identical to the period-by-period means when the number of sales is the same for all periods in the repeat sales sample. Thus, McMillen (2012) suggests taking the propensity score matching approach to produce a matched sample; this approach pairs sales in a base period with similar properties selling at other times. With a large sample size, averaging matched sales prices at each time point is directly comparable to the repeat sales estimator. A hedonic approach can be used with a matched sample, and this makes the estimates less model-dependent.

Mean-based price indices reflect only one aspect of the distribution. Unlike mean-based price indices—which focus on mean-price changes—McMillen (2012) suggests using a quantile approach to estimate an index for any point in the price distribution, such as the median, the 10th percentile, or the 90th percentile. As

[6] Shimizu et al. (2010) analyzed the lead-lag structure of hedonic price indices, BMN-type repeat sales indices, Case–Shiller-type repeat sales indices, and age-adjusted repeat sales indices. Their results clearly show that even when adjusted for age, repeat sales indices lag relative to hedonic price indices. This makes it evident that when it comes to the causes of this lag, sample selection bias is more significant than the lack of age adjustment.

Koenker and Bassett (1978) state—and which was subsequently modified by McMillen (forthcoming)—the quantile regression formation can be described as:

$$Q_{lnP}(q|X_{it}, D_{it}) = X_i\beta(q) + \sum_{t=2}^{T} D_{it}\delta_t(q) \quad (3)$$

Here, Q represents the conditional natural log of the sale price at quantile $q \in (0,1)$, X is the combination of variables that represent the structure and the location characteristics, and D is the matrix of the time dummy. This formation can trace out the full distribution by estimating across different quantiles.

The aforementioned discussion pertains to changes in price distribution. Another stream of literature focuses on approaches that allow for spatial variation in appreciation rates. Cleveland and Devlin (1988) introduced the locally weighted regression approach, in which the coefficients of the estimating equation are assumed to vary smoothly over space. McMillen (forthcoming) constructed a local quantile house price index based on *loess* and quantile regressions, and it allows spatial variations in the appreciation rate. In the local quantile index, locally weighted models are estimated for a set of target locations, with more weight being assigned to observations that are closer to the target locations. Compared to the quantile model that includes neighborhood fixed effects, the local weighted approach has advantages. First, it is more convincing to assume that the spatial effects vary smoothly, rather than change discretely at neighborhood boundaries. Second, having within one's sample only a few observations in small locations will lead to imprecise estimates, and it can lead to problems of convergence for the quantile estimator.

In this study, we will use the locally weighted quantile approach to investigate how the distribution of price in the Greater Tokyo and Kansai areas changed in the 1986–2015 period. It is important to compare the changes in house price distribution in these two areas, which constitute Japan's largest metropolitan areas. As each metropolitan area includes many cities and the appreciation rates within each metropolitan area may vary over space, the use here of the locally weighted quantile approach is sound.

3. Data

The dataset used in this study contains condominium data listings for two large Japanese metropolitan areas (i.e., the Greater Tokyo and Kansai areas) over 30 years (i.e., 120 quarters), starting from the first quarter of 1986 (i.e., 1986Q1) and ending with the fourth quarter of 2015 (i.e., 2015Q4). This dataset is provided by Suumo (Residential Information Website), which is owned by Recruit Co., Ltd., one of the largest vendors of residential lettings information in Japan.[7] This dataset contains the final week's listing price, just before removal due to sale.[8] The structural characteristics include floor space and building age. The amenity characteristics

[7] Shimizu et al. (2004) report that the Recruit data cover more than 95% of all transactions in Tokyo's 23 special wards. On the other hand, its coverage for suburban areas is very limited. We therefore use only information for the units located in the special wards of Tokyo.

[8] There are two reasons for the removal of a unit's listing from the magazine: a successful sale, or a withdrawal (i.e., the seller gives up looking for a buyer and thus withdraws the listing). We were allowed to access information regarding which of the two reasons applied to each individual case, and discarded those where the seller withdrew the listing. Shimizu et al. (2016) show that in this dataset, the final week's listing price is almost identical to the contract price.

include walking time to the nearest station, time by train in the day time to the nearest terminal station, and to the city center. The location characteristics include city code, address, latitude, and longitude. There are 438,020 observations for the Greater Tokyo area, and 119,779 observations for the Kansai area.

In this dataset, house transactions in the Greater Tokyo area involve three large cities (Tokyo, Yokohama, and Kawasaki) and several small cities (Saitama, Chiba, Hachioji, Mitaka, and Sagmihara). House transactions in the Kansai area involve three large cities (Osaka, Kyoto, and Kobe) and three small cities (Otsu, Sakai, and Nara). The transactions in Tokyo's 23 special wards account for 44% of all transactions, while those in Osaka account for 10% of all transactions.

The first and third panels of Table 1 present summary statistics of the full sample of the Greater Tokyo and Kansai areas. The average listing price in the Greater Tokyo area in 1986 was 25 million JPY; in 1990, it was 55 million JPY. This reveals that house prices peaked in 1990, followed by a burst period that lasted until 2000, at which point the average price was 28 million JPY. The average listing price in the Kansai area was 20 million JPY in both 1986 and 2000; in 1990, at the peak of the boom, it was 54 million JPY. The difference between the average price in Greater Tokyo and that of Kansai became large in 2005.

The average characteristics of the houses changed over time. Before the burst, properties tended to be newer and smaller. Both size and building age increased between 1986 and 2000. In each year in the dataset, the average floor space in the

Kansai area is larger than that in the Greater Tokyo area. While the average building age in the Kansai area between 1986 and 2000 tends to be smaller than that in the Greater Tokyo area, after 2005, they converge to become more similar. Commuting times are small for houses sold in 1986, in both the Greater Tokyo and Kansai areas.

As discussed previously, McMillen (2012) points out that 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. The use of a matching procedure to pre-process the data helps make the estimates less model-dependent. It is important to recognize that the initial matching approach does not include controls for location (i.e., city code or geographic coordinates). McMillen (forthcoming) states that controlling for location in the initial matching procedure is problematic, since including neighborhood indicator variables could lead to small samples within each neighborhood over time.

In this study, we first use propensity score matching to prepare a matched sample. Using 2000Q1 as the base each time, we estimate a series of logit models for each subsequent quarter. First, for each quarter q from 1986Q1 to 2015Q4 (excluding 2000Q1), we estimate a logit regression of all sales in 2000Q1 and q. The dependent variable equals 1 if the sale took place in q, and 0 if the sale took place in 2000Q1. The explanatory variables are identical to those used in the hedonic regressions (except geographic coordinates): these include floor space, building age, time to nearest station, time to terminal, and time to CBD. Second, we use the estimated propensity score from each logit regression to match sales from quarter q to its closest counterpart in 2000Q1. Whenever the number of observations in quarter q exceeds that in 2000Q1, the matched sample for quarter q will contain no more than the number of observations in 2000Q1, n_0. In the quarters where the number of observations in quarter q is less than n_0, all observations will remain in the matched sample, unless this support condition is violated. We estimate this matching approach for the Greater Tokyo and Kansai areas; we provide summary statistics in the second and fourth panels of Table 1.

In the matched sample, the average price in the Greater Tokyo area in 1990 is 60 million JPY; this amount is larger than the 55 million JPY average for the full sample. Meanwhile, for the Greater Tokyo area, the average price for the matched sample in other years is similar to that for the full sample. It is noted that the difference in unit price between the matched sample and the full sample is small, since the floor space in the matched sample is similar for each period. Over time, in the matched sample, the variation in the means of floor space and building age are quite small. This finding aligns with the assertions of McMillen (forthcoming). In the Empirical Results section, all results are based on the matched sample.

4. Methodology

In this section, we present a locally weighted quantile regression method that is used to estimate the set of matched samples for the Greater Tokyo and Kansai areas.

To start, we estimate locally weighted models for a set of target locations, with more weighted being assigned 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 assigned to observation i when estimating the model for the target location is $K(d_i)$, where K is any standard kernel weight function. For linear regression models, weighted least square regression is used as the estimation procedure for each target location. For the quantile approach, the weighted version can be expressed as finding the $\hat{\beta}(q,d)$ that minimizes the locally weighted objective function $\sum_i K(d_i)\rho_q(y_i - x_i'\beta)$. As is the case with the standard geographically weighted regression first used by McMillen (1996), the thinking behind this quantile approach is to assign more weight to nearby observations when estimating the model at the target point.

In many studies, each observation serves in turn as a separate target point. However, the estimation time can be reduced significantly by taking advantage of the smoothness implied by the *loess* approach, by interpolating from a smaller set of target points to each location represented in the dataset. In this case, we estimate the models separately for the Greater Tokyo and Kansai areas. The quantile range of $q =$ 0.02–0.98, in increments of 0.02, implies a number of quantiles $B = 48$. We estimate the quantile regressions at a set of target locations and then interpolate to all other locations in the dataset for each region. We use a tri-cube kernel with a 25% window, based on the straight-line distance between each observation and the target point. The result of the estimation procedure is a set of $n \times B \times k$ estimated coefficients for each metropolitan area, as we know the explanatory variables X and sale date for each observation. (n is number of observations for each region and k is number of explanatory variables.) We predict a set of $n \times B$ counterfactual prices for each metropolitan area, by setting X to its actual value while setting $D_t = 1$ for the year under consideration and the values of the dummy variables to 0 for all other years. In this way, the estimated density in each year for each of the Greater Tokyo and Kansai areas can be easily derived.

As the values of the explanatory variables changed considerably during the bubble period, it is important to simulate the change in the estimated price distribution by controlling for a variable at a designed value, while setting all other variables to their actual values. Besides, we can trace the effect of a change in a structural characteristic by comparing the estimated densities at two or more values, while other variables remain as actual values. For any explanatory variable j, we calculate kernel density estimates for $n \times B$ estimated values of $z_j\beta_j(q) + \sum_{k \neq j} x_k\beta_k(q) + D_t\delta_t(q)$, where z_j is a set of designed values.

5. Empirical Results

5.1 Hedonic Price Indices

We first conduct a standard hedonic regression for the log of prices for matched samples of the Greater Tokyo and Kansai areas, using a specification similar to that adopted by Shimizu et al. (2010, 2016), among others. The logarithmic house price is the dependent variable, while the explanatory variables are structural features (e.g., floor space and building age) and location characteristics (e.g., commuting time to station, terminal station, and city center). After controlling for latitude, longitude, and quarterly time dummies, the regression equation is as follows:

$$\mathrm{Ln}(Price) = \beta_0 + \beta_1 area + \beta_2 age + \beta_3 t_{station} + \beta_4 t_{center} \quad (6)$$
$$+\beta_5 t_{CBD} + \beta_6 latitude + \beta_7 logtitude + D_{year}\beta \ + \epsilon$$

The regression results are shown in the Table 2. The results are standard for both the Greater Tokyo and Kansai areas: house prices increase with floor space and decline with building age; in addition, prices decline with greater commuting time to nearest station, terminal station, and the CBD.

To accurately examine the timing of the boom and of the burst, we also estimate the hedonic model while controlling for the quarterly time dummies. The quarterly hedonic price indices for the Greater Tokyo and Kansai areas are shown in Figure 1. For both areas, the peak of the boom is in 1990; however, the beginning of the boom in the Greater Tokyo area occurred in 1986, which is somewhat earlier than that in the Kansai area (i.e., after 1987). The appreciation rate for the Kansai area was incredibly large during the 1987–1990 period. For the Kansai area, the postpeak decline occurred a little earlier than did that of the Greater Tokyo area; additionally, the burst in the Greater Tokyo area stopped in 2000—even as house prices in the Kansai area continued to decline, until 2003. In both areas, house prices recovered after 2006 and declined again in the 2008–2010 period, principally because of the subprime crisis in the United States. House prices again recovered after 2012, to reach their highest postbubble level.

5.2 Price Percentiles of Matched sample

Hedonic price indices cannot be compared directly, since a large number of explanatory variables are being controlled for. In this section, we simply show how the 10th, 50th, and 90th percentiles of the price distribution vary over time for two metropolitan areas. Figures 2 and 3 show how, for the matched sample, there was a change over time in the percentiles of the price distribution.

Figure 2 compares the various percentiles within the two metropolitan areas. In the boom of the Greater Tokyo area, house prices saw a temporary decline in each percentile in the 1987–1989 period. The prices of high-priced houses declined more so than those of middle-priced houses, while the prices of low-priced houses remained stable during this period. There was no temporary decline in the Kansai area during this period. We find that the timing of the boom and of the burst varied across the percentiles. The start of the boom for high-priced houses was earlier than that of low-priced houses, in both the Greater Tokyo and Kansai areas. After 1990, the start of the burst for high-priced houses occurred later than that for middle and low-priced houses, in both the Greater Tokyo and Kansai areas.

To compare the two metropolitan areas, Figure 3 shows the price changes for the 10th, 50th, and 90th percentiles. Overall, during the bubble period, house prices in Greater Tokyo appreciated earlier and declined later than did those in the Kansai area. Before 2000, the difference in price between the two metropolitan areas was large for high-priced houses and small for low-priced houses. House prices in the Kansai area continued to decline in 2000–2005, while those in the Greater Tokyo area remained stable.

5.3 Locally Weighted Quantile Estimates

In this section, we present the results of locally weighted quantile regression estimates for the set of matched samples. As mentioned in section 4, we take the locally weighted quantile approach to generate a counterfactual distribution for each year, for each metropolitan area. Figure 4 presents the estimated distributions of two sets of years. We select three years—namely, 1986, 1990, and 2000—to show the change in distribution in the bubble period, and the change in three recent years (i.e., 2005, 2010, and 2015). These densities indicate that not only the price changed over time, but also that variance changed considerably. The prices of the middle-priced houses were similar for 1986 and 2000 in the Kansai area, but the variances were very different. The variance of house prices in the Kansai area 1986 was extremely small. Considering that the average building age in the Kansai area was quite small in 1986, we believe that many new middle-priced houses were constructed in that area before the bubble occurred.

Figure 5 presents violin plots, to show the estimated distributions. The blue violins are estimated densities for the Greater Tokyo area, and the green violins are estimated densities for the Kansai area. We select seven years to show the change in densities. The density of the Kansai area in 1986 shows small variance and many middle-priced houses, while the variance of density of the Kansai area in 1900 is smaller than that of Greater Tokyo. The difference in price increased after 2000, as did the difference in variance. After 2000, the variance in house prices in Kansai increased, and the depreciation rate of low-priced houses was larger than those of middle and high-priced houses.

5.4 Counterfactual Distribution

In this section, we calculate counterfactual distributions by simulating the values for variables set at representative values: the floor space is set to 50, 65, or 85 m^2; building age is set to 8, 13, or 19 years; and the time to the CBD is set to 15, 30, or 45 mins.[9]

Figure 6 shows the counterfactual distributions for the observations of every year. Panels A and B show how changes in floor space impact the distributions. As expected, increases in floor space shift the sale price distribution rightwards. The variance declines as floor space increases. When controlling for the floor space, the distributions have fat left tails in the Kansai area and fat right tails in the Greater Tokyo area. The median values and variance of the sale price distribution decrease

[9] For each variable, a set of values is selected from values close to 25%, 50%, and 75%.

with housing age. The shifts of distributions of building age (panels C and D) are larger in Kansai than in Greater Tokyo. Additionally, the distributions shift leftwards and the variance increases with a larger commuting time to the CBD (panels E and F). The shifts in distribution with a larger commuting time to the CBD are larger in Greater Tokyo than in Kansai.

In Figure 7, we present the changes in counterfactual distributions in 1986, 1990, and 2000. In panels A and B, we set the floor space at 65 m^2; in panels C and D, we set the building age at 13 years; and in panels E and F, we set the time to the CBD at 30 min. If we compare the distributions of the Greater Tokyo area to those in Figure 4, we find that neither the building age nor the commuting time to the CBD overly dictates the distribution. The floor space setting shifted the 1986 distribution rightwards; this may be due to the average floor space in Greater Tokyo in 1986 being 51.42 m^2, whereas the overall average was 61.49 m^2. Many small units were on the market before the boom. Unlike with the Greater Tokyo area, in the Kansai area, the variance increased considerably with house prices in 1986, while controlling for floor space. When we set the building age to 13 years, both the price and variance decreased in Kansai in 1986, and this suggests that new houses exacerbated the variance.

5.4 Local Appreciation Rate

In this section, we show the local appreciation rates of the two metropolitan areas. Figure 8 presents the appreciation rates for 1986–1990 and 1990–2000, for small, local districts. In that figure, wherever there is more red (panels A and B) and dark blue (panels C and D), it means the absolute values of the appreciation rate are large. During the bubble and the burst, the absolute values of the appreciation rates were large in Osaka, Kyoto, and Nara. However, the northern part of Kobe had the largest appreciation rate in the 1986–1990 period, while the southern part saw a greater decline in the 1990–2000 period. The 23 special wards of the Greater Tokyo area saw the biggest decline of the 1990–2000 period. The regions that saw the largest appreciation rates are not limited to those in special wards.

6. Conclusion

This study proposed a price indices estimation method that makes it possible to estimate temporal changes in the price distribution of large metropolitan areas in Japan, by applying a locally weighted quantile regression. Specifically, it developed an index that makes it possible not only to estimate the typical median-based price indices, but also to capture temporal changes in different market segments (i.e., low-priced properties versus high-priced properties) in large metropolitan areas, by generating a quality-adjusted price distribution for each period; this is done by using quantile regression.

Consequently, compared to the standard hedonic price method—which is the estimation method typically used for price indices—the following results were obtained.

For house price indices estimated for the Greater Tokyo and Kansai areas:

- Looking at the period during which the bubble formed in Japan during the 1980s, the price index for high-priced properties (i.e., 90th percentile) increased ahead of the other indices.
- On the other hand, when the bubble started to collapse in 1990, the decline began with the price index for low-priced properties (i.e., 10th percentile).

These kinds of differences in price level and distribution are estimated in terms of price distribution differences, using results that reflect spatial differences (i.e., upscale residential neighborhoods, standard residential neighborhoods, and disadvantaged residential neighborhoods), as well as differences in grade (i.e., luxury housing, standard housing, and low-quality housing). In other words, estimating the price distribution makes it possible to construct a price index without performing spatial segmentation or quality segmentation. Moreover, the fact that certain price ranges were identified as preceding or lagging behind the overall market suggests that price index values derived in this way offer more information than price index values that are based on the standard hedonic method or the repeat sales method.

In addition, based on a comparison of the distributions of the two metropolitan areas:

- The 1980s' bubble occurred between 1986 and 1987, inclusive, in the Greater Tokyo; then, just as the bubble was almost coming to an end, price increases occurred in the Kansai area. This finding suggests that the bubble had spread from the Greater Tokyo area to the Kansai area in 1987. Subsequently, although the original Tokyo housing bubble had almost come to an end, prices in Tokyo began to rise once again, starting in 1988, influenced by the rising prices in the Kansai region; the bubble then collapsed, in 1990.

- Our estimates show that the bubble collapsed abruptly in the Kansai area, and after that time, prices gradually began to decline in the Greater Tokyo area.

As mentioned previously, when using conventional hedonic price indices and repeat sales indices, it was not possible to anticipate the formation and collapse of the aforementioned Japanese property bubble.

Generally, the purpose of a price index is to capture average fluctuations within a given market; however, since the housing market possesses a high degree of heterogeneity in terms of space and grade, there are many cases in which it is important to observe changes in more segmented market units. Furthermore, when it comes to comparing housing purchasing power, it is necessary to observe not just price changes, but also price levels. The importance of this can be seen, for example, in the government policy targets set during the bubble era, with the purpose of controlling house prices so as not to exceed annual income by more than five-fold.

Accordingly, it is possible to develop an affordability index by comparing the average income of each city, and a quality-adjusted house price distribution and index by using quantile regression for each city examined here.

Furthermore, the method proposed by this study enables one to estimate price fluctuations and price levels by city in a simultaneous manner; additionally, it leverages data readily obtained online, and this is another benefit.

Following the publication of the *International Handbook on Residential Property Prices*, statistical agencies in various countries have been trying to move forward in developing house price indices. However, there are various issues at hand, such as data limitations, and progress being made in developing these indices is dependent on the data sources and methods being used. If it were possible to obtain international house price information online, it would be possible to estimate indices with identical methods, using equivalent data sources; this would facilitate the development of house price indices and affordability indices that allow for truly meaningful international comparisons. We intend to address this issue, among other remaining research topics.

References

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association, 58*(304), 933-942.

Case, B., & Quigley, J. M. (1991). The Dynamics of Real Estate Prices. *The Review of Economics and Statistics, 73*(1), 50-58.

Case, K. E., & Shiller, R. J. (1989). The Efficiency of the Market for Single-Family Homes. *The American Economic Review, 79*(1), 125-137.

Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association, 83*(403), 596-610.

Deng, Y., McMillen, D. P., & Sing, T. F. (2012). Private residential price indices in Singapore: A matching approach. *Regional Science and Urban Economics, 42*(3), 485-494.

Deng, Y., McMillen, D. P., & Sing, T. F. (2014). Matching indices for thinly-traded commercial real estate in Singapore. *Regional Science and Urban Economics, 47*, 86-98.

Diewert, E., & Shimizu, C. (2015). RESIDENTIAL PROPERTY PRICE INDICES FOR TOKYO. *Macroeconomic Dynamics, 19*(8), 1659-1714.

Diewert, W. E., & Shimizu, C. (2016). Hedonic regression models for Tokyo condominium sales. *Regional Science and Urban Economics, 60*, 300-315.

Karato, K., Movshuk, O., & Shimizu, C. (2015). Semiparametric Model of Hedonic Housing Prices in Japan. *Asian Economic Journal, 29*(4), 325-345.

Koenker, R. (2005). *Quantile regression* (Vol. no. 38.). Cambridge: Cambridge University Press.

Koenker, R., & Jr, G. B. (1978). REGRESSION QUANTILES. *Econometrica (pre-1986), 46*(1), 33-50.

Mark, J. H., & Goldberg, M. A. (1984). Alternative Housing Price Indices: An Evaluation. *Real Estate Economics, 12*(1), 30-49.

McMillen, D. P. (1996). One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. *Journal of Urban Economics, 40*(1), 100-124.

McMillen, D. P. (2008). Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients? *Journal of Urban Economics, 64*(3), 573-589.

McMillen, D. P. (2012). Repeat Sales as a Matching Estimator. *Real Estate Economics, 40*(4), 743-771.

McMillen, D. P. (2013). *Quantile regression for spatial data*. New York;Berlin;: Springer.

McMillen, D. P. (forthcoming). Local quantile house price indices. *Journal of Urban Economics*.

Palmquist, R. B. (1980). Alternative Techniques for Developing Real Estate Price Indexes. *The Review of Economics and Statistics, 62*(3), 442-448.

Rosen, S. (1974). HEDONIC PRICES AND IMPLICIT MARKETS - PRODUCT DIFFERENTIATION IN PURE COMPETITION. *The Journal of Political Economy, 82*(1), 34-55.

Shimizu, C., Nishimura, K. G., & Asami, Y. (2004). SEARCH AND VACANCY COSTS IN THE TOKYO HOUSING MARKET: ATTEMPT TO MEASURE SOCIAL COSTS OF IMPERFECT INFORMATION. *Review of Urban & Regional Development Studies, 16*(3), 210-230.

Shimizu, C., Nishimura, K. G., & Watanabe, T. (2010). Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics, 230*(6), 792-813.

Shimizu, C., Nishimura, K. G., & Watanabe, T. (2016). House prices at different stages of the buying/selling process. *Regional Science and Urban Economics, 59*, 37-53.

Shimizu, C., Takatsuji, H., Ono, H., & Nishimura, K. G. (2010). Structural and temporal changes in the housing market and hedonic housing price indices. *International Journal of Housing Markets and Analysis, 3*(4), 351-368.

Wang, F. T., & Zorn, P. M. (1997). Estimating House Price Growth with Repeat Sales Data: What's the Aim of the Game? *Journal of Housing Economics, 6*(2), 93-118.

Table 1. Means of Full and Matched Samples

	Total	1986	1990	1995	2000	2005	2010	2015
				Greater Tokyo (Full)				
List Price	3668.6	2522.0	5500.2	3529.5	2809.4	2996.1	3220.9	3632.3
Unit Price	61.69	51.75	104.3	57.85	43.64	45.76	50.04	56.60
Floor Space	61.49	51.42	55.59	61.08	64.08	65.45	65.30	65.39
Building Age	163.2	95.25	119.9	160.3	176.1	169.0	186.6	212.6
Time to Station	8.500	7.622	8.792	8.681	8.331	7.992	8.396	8.627
Time to Terminal	17.06	13.35	17.94	17.98	17.57	16.01	16.59	16.51
Time to CBD	34.64	29.93	35.65	35.85	35.33	33.15	34.57	33.67
N	438020	6853	17493	26642	8122	7533	7149	18343
				Greater Tokyo (Matched)				
List Price	3662.9	2539.7	6034.6	3772.2	2800.5	2988.0	3216.7	3624.9
Unit Price	58.96	51.03	100.8	58.05	43.56	45.71	50.02	58.37
Floor Space	63.17	52.11	61.39	64.14	64.02	65.39	65.26	64.05
Building Age	165.9	96.79	151.2	173.2	176.1	169.1	186.5	175.5
Time to Station	8.359	7.681	8.610	8.397	8.328	7.993	8.399	8.395
Time to Terminal	16.97	13.51	17.93	17.84	17.61	16.02	16.59	17.49
Time to CBD	34.69	30.11	36.04	35.77	35.41	33.16	34.58	35.21
N	235048	6704	8418	8501	8038	7527	7145	8462
				Kansai (Full)				
List Price	2783.7	2025.3	5444.5	2968.8	2016.9	1725.9	2067.9	2277.9
Unit Price	41.96	31.09	90.71	44.05	28.55	24.54	29.17	33.01
Floor Space	67.85	65.20	61.75	67.39	69.92	69.82	70.57	69.22
Building Age	157.0	78.14	108.5	141.6	173.2	180.8	189.5	216.6
Time to Station	7.553	7.202	7.647	7.951	7.615	7.706	7.140	7.076
Time to Terminal	14.97	11.35	14.73	15.71	16.31	15.83	14.86	13.59
Time to CBD	32.73	27.90	29.44	32.19	33.11	39.25	35.21	31.25
N	119779	336	6287	6567	1924	2033	2859	5832
				Kansai (Matched)				
List Price	2499.5	2041.0	5515.4	2928.7	2013.8	1714.5	2033.4	2303.7
Unit Price	36.12	30.96	80.78	42.15	28.54	24.38	28.92	33.60
Floor Space	69.15	65.79	68.02	69.06	69.86	69.89	70.07	69.13
Building Age	166.9	79.06	149.3	169.5	173.0	179.6	177.4	171.6
Time to Station	7.728	7.146	7.762	7.957	7.624	7.781	7.365	7.583
Time to Terminal	15.91	11.49	16.42	16.29	16.33	15.97	15.00	16.10
Time to CBD	33.79	28.02	33.21	33.53	33.07	38.86	31.59	32.61
N	57883	329	2064	2119	1915	1964	2119	2126

Notes: The first and third panels are full samples of the Greater Tokyo and Kansai areas; the second and fourth panels are matched samples of the Greater Tokyo and Kansai areas. Listing price is the final week's listing price (10,000s of JPY). Unit price is the price per square meter (10,000s of JPY/m^2). Floor space is the floor area (m^2), building age is the age of the building between the date of construction and the transaction (months), time to nearest station is the walking time to the nearest station (minutes, with 1 min equaling approximately 80 m), and time to terminal station is the commuting time in taking the train from the nearest station to the nearest terminal station in the daytime (minutes). Time to CBD is the commuting time in taking the train from the nearest station to the central business district (i.e., Tokyo Station in Greater Tokyo, or Umeda Station in Kansai) in the daytime (minutes). N is number of the observations.

Table 2. Hedonic Regression Results

	Greater Tokyo	Kansai
Floor Space	0.017***	0.016***
	(0.00003)	(0.0001)
Building Age	-0.001***	-0.002***
	(0.00001)	(0.00001)
Time to Station	-0.019***	-0.015***
	(0.0001)	(0.0002)
Time to Terminal	-0.006***	-0.008***
	(0.0001)	(0.0002)
Time to CBD	-0.012***	-0.004***
	(0.0001)	(0.0001)
Year Fixed	Y	Y
Geographic Coordinates Fixed	Y	Y
Constant	90.144***	-2.455***
	(0.934)	(0.587)
R^2	0.724	0.797
N	235,048	57,883

Notes: The dependent variable is log of house price. Robust standard errors are in parentheses.

* p<0.1, ** p<0.5, *** p<0.01.

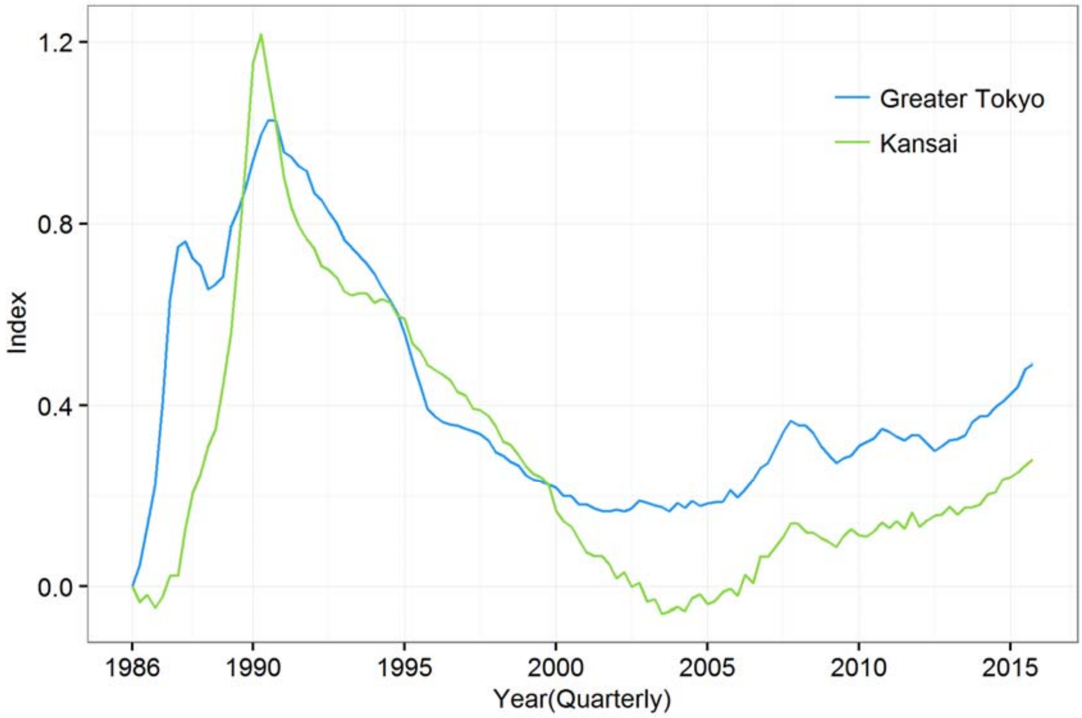
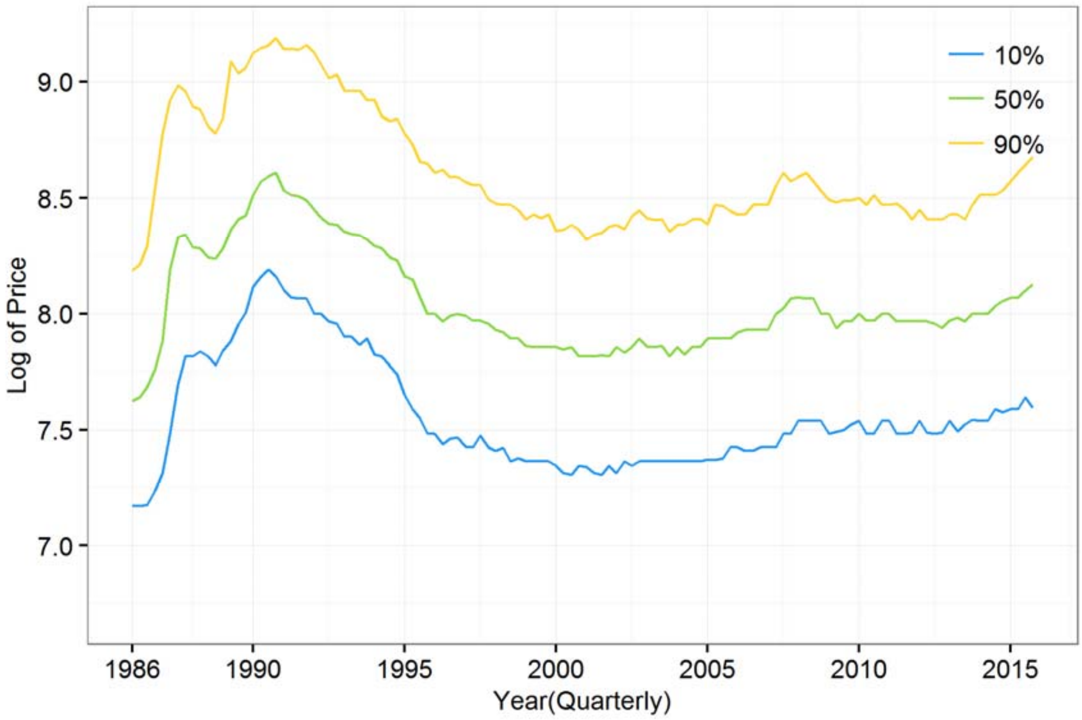
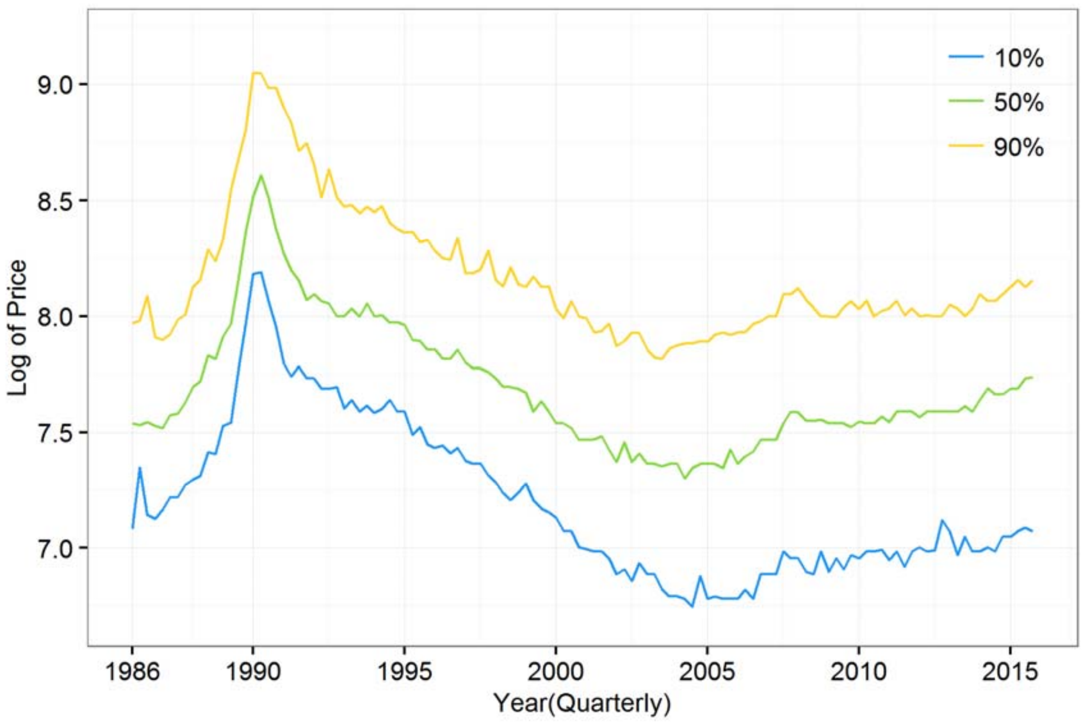


Figure 1. Hedonic Price Indices

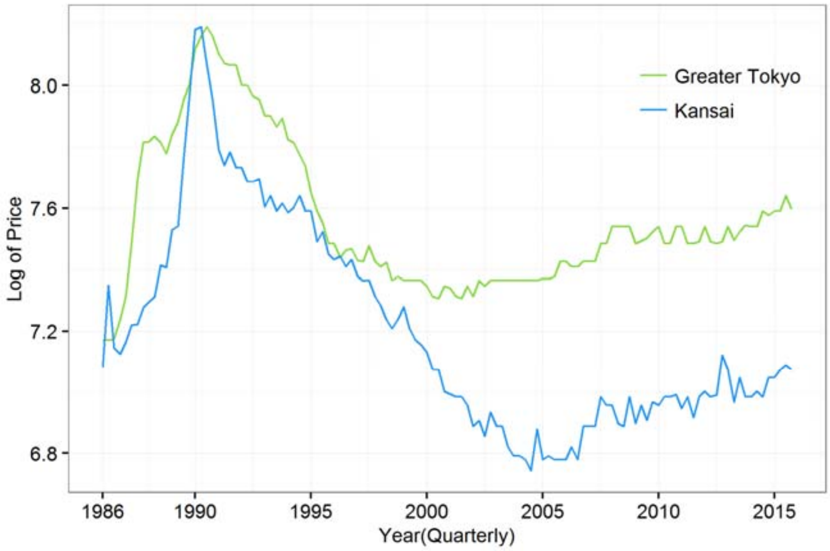


A. Greater Tokyo

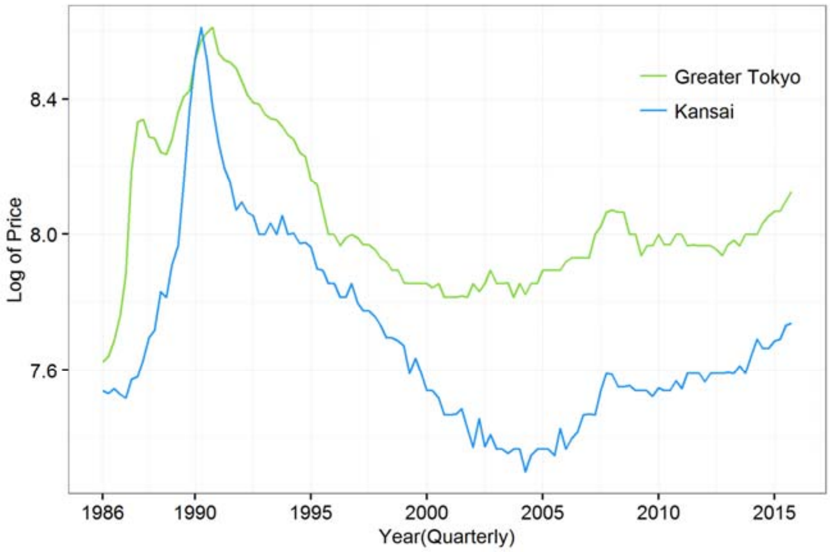


B. Kansai

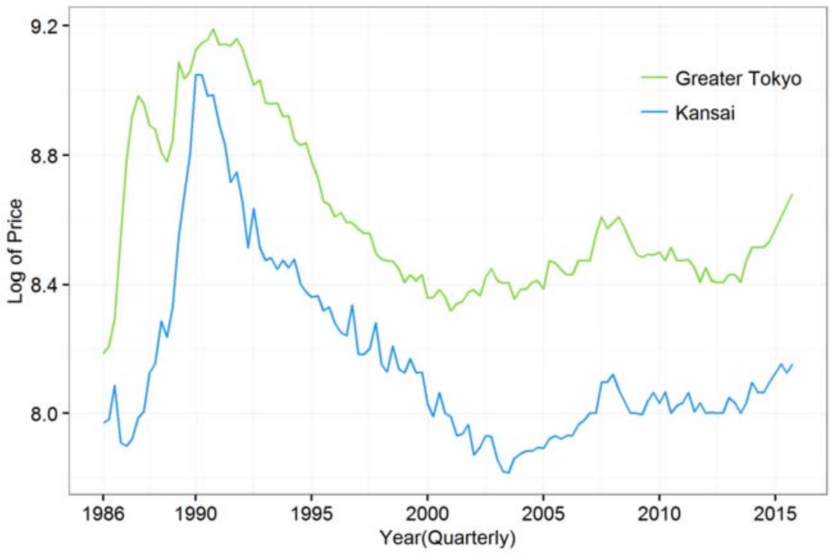
Figure 2. Sale Price Percentiles of Matched Sample, by Year



A. 10th Percentile

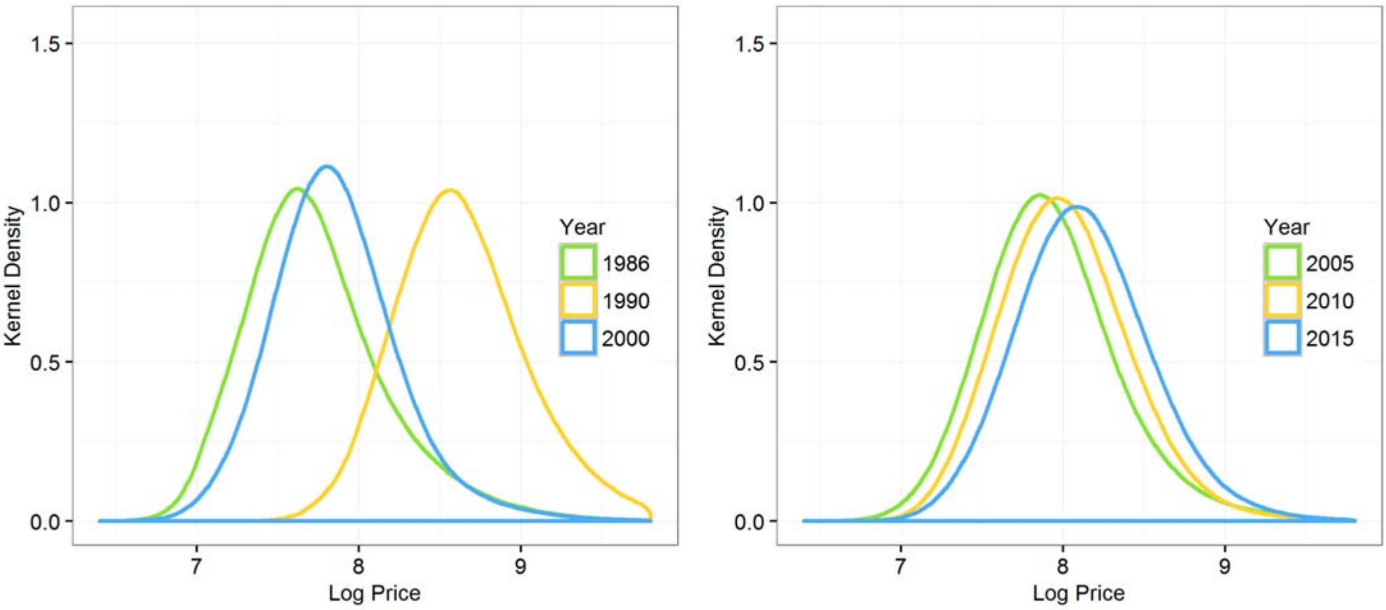


B. 50th Percentile

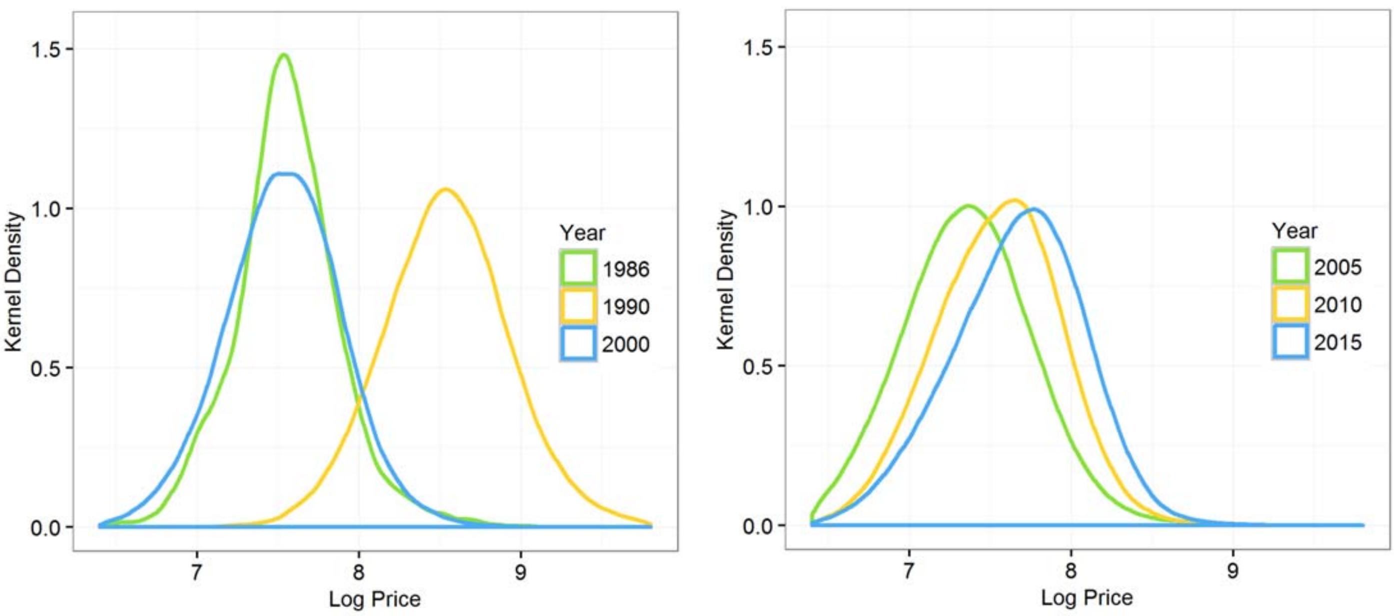


C. 90th Percentile

Figure 3. Sale Price Percentile of Matched Sample, by City



A. Greater Tokyo



B. Kansai

Figure 4: Estimated Density, by Local Quantile Regressions

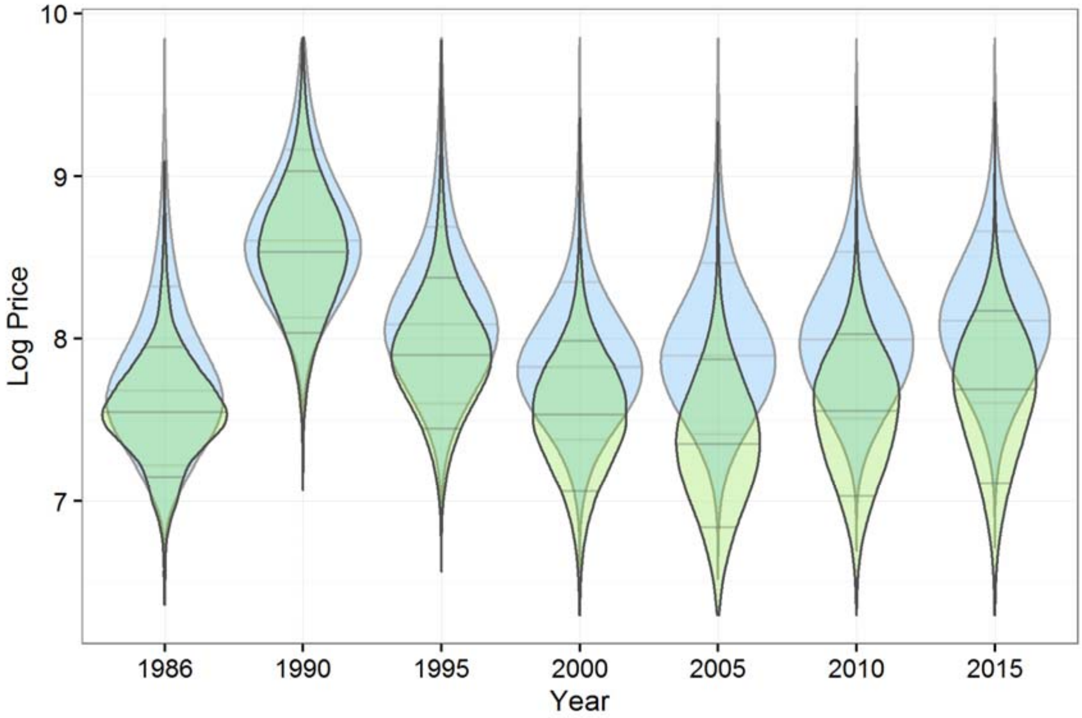
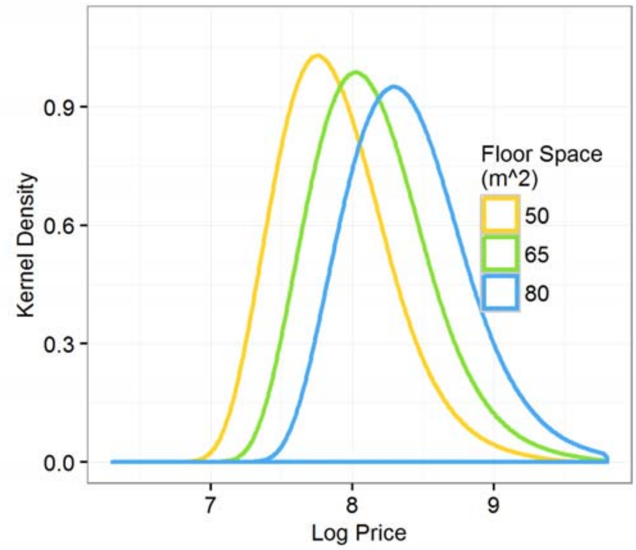
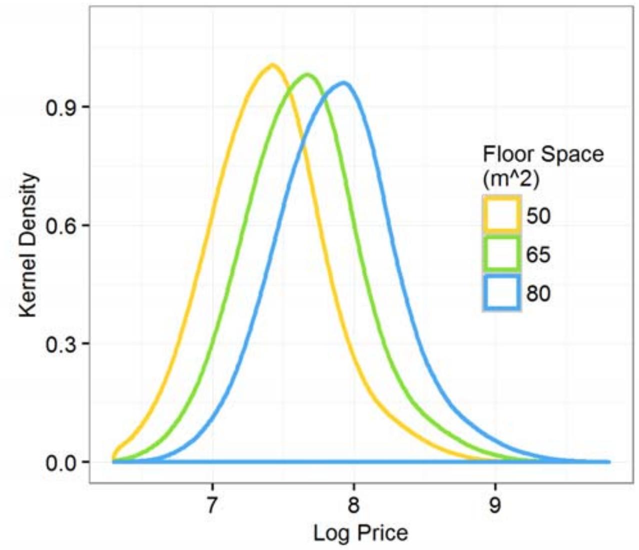


Figure 5. Violin Plots of Estimated Density

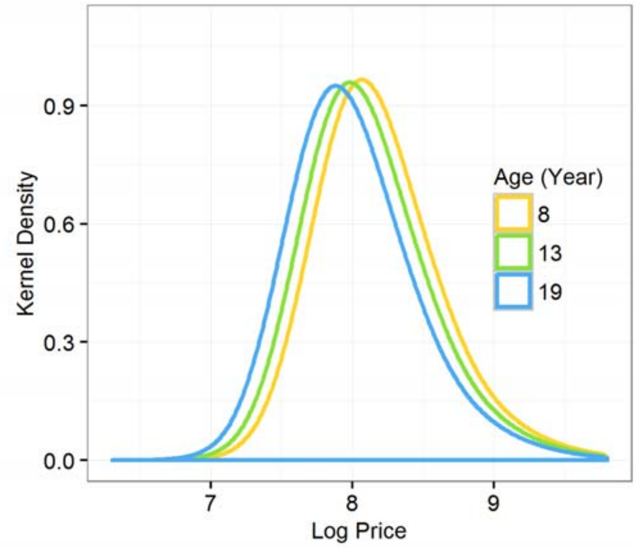
Note: Blue violins represent the estimated density of the Greater Tokyo area; green violins represent the estimated density of the Kansai area. The three lines in each violin represent prices in the 10th, 50th, and 90th percentiles.



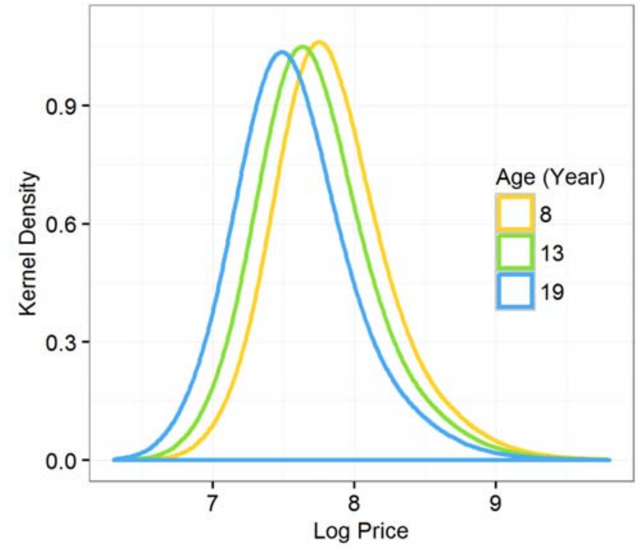
A. Floor Space: Greater Tokyo



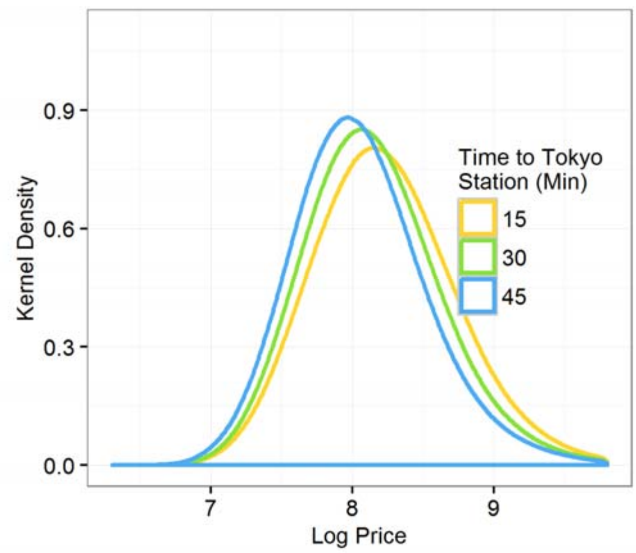
B. Floor Space: Kansai



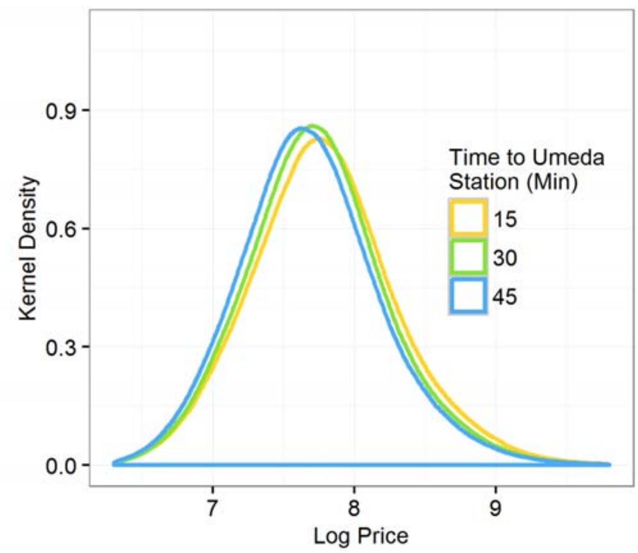
C. Building Age: Greater Tokyo



D. Building Age: Kansai

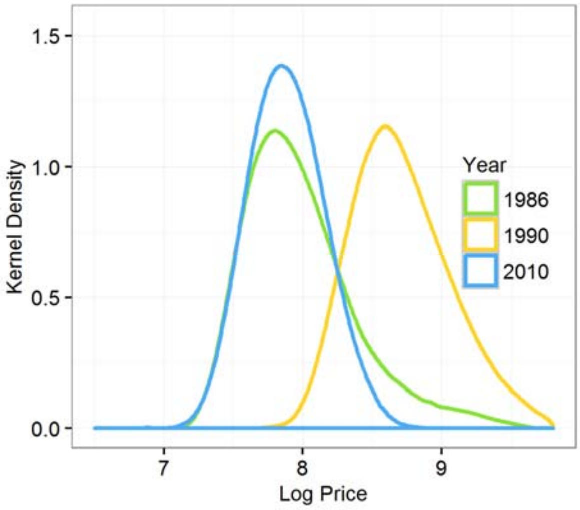


E. Time to CBD: Greater Tokyo

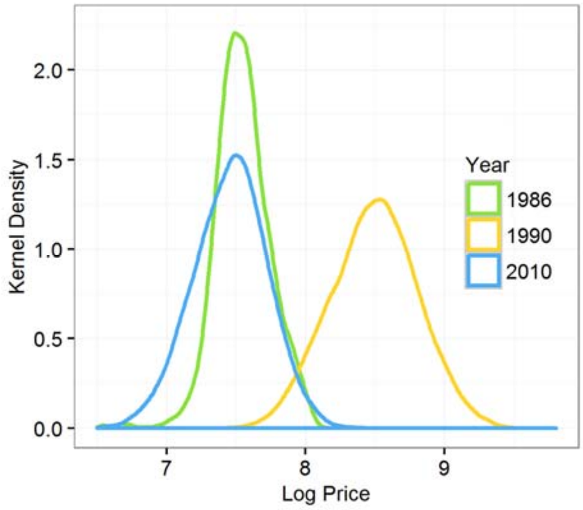


F. Time to CBD: Kansai

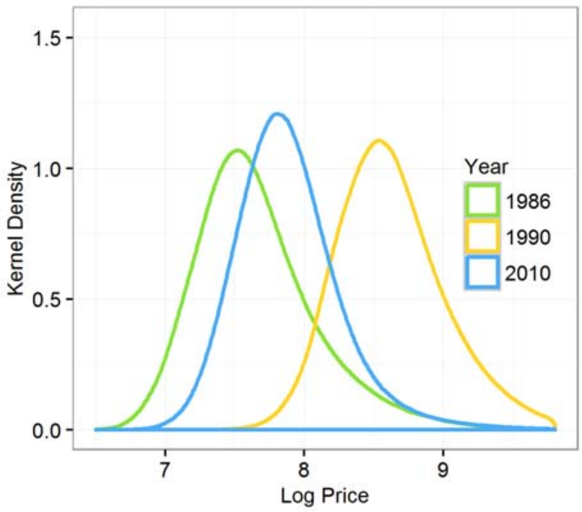
Figure 6. Estimated Counterfactual Density



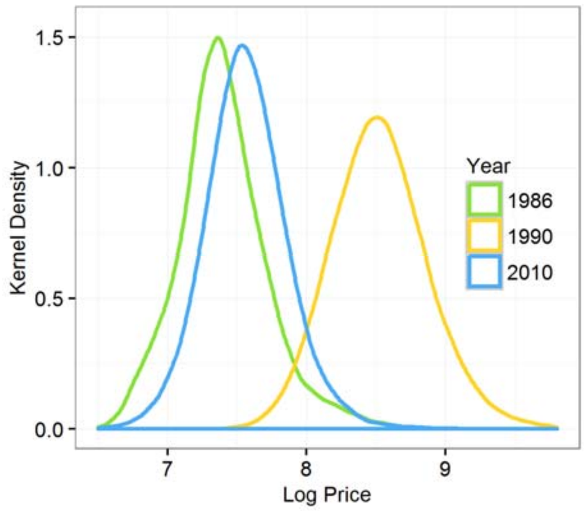
A. Floor Space (65 m^2): Greater Tokyo



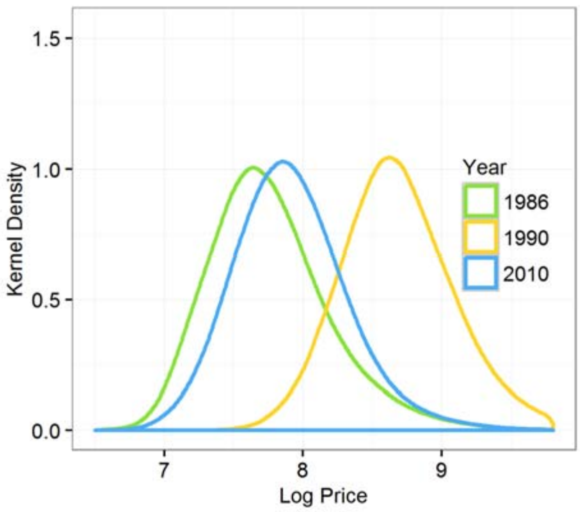
B. Floor Space (65 m^2): Kansai



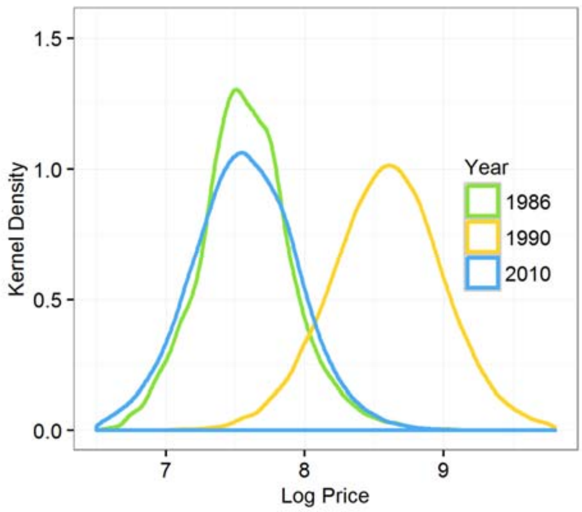
C. Building Age (13 years): Greater Tokyo



D. Building Age (13 years): Kansai

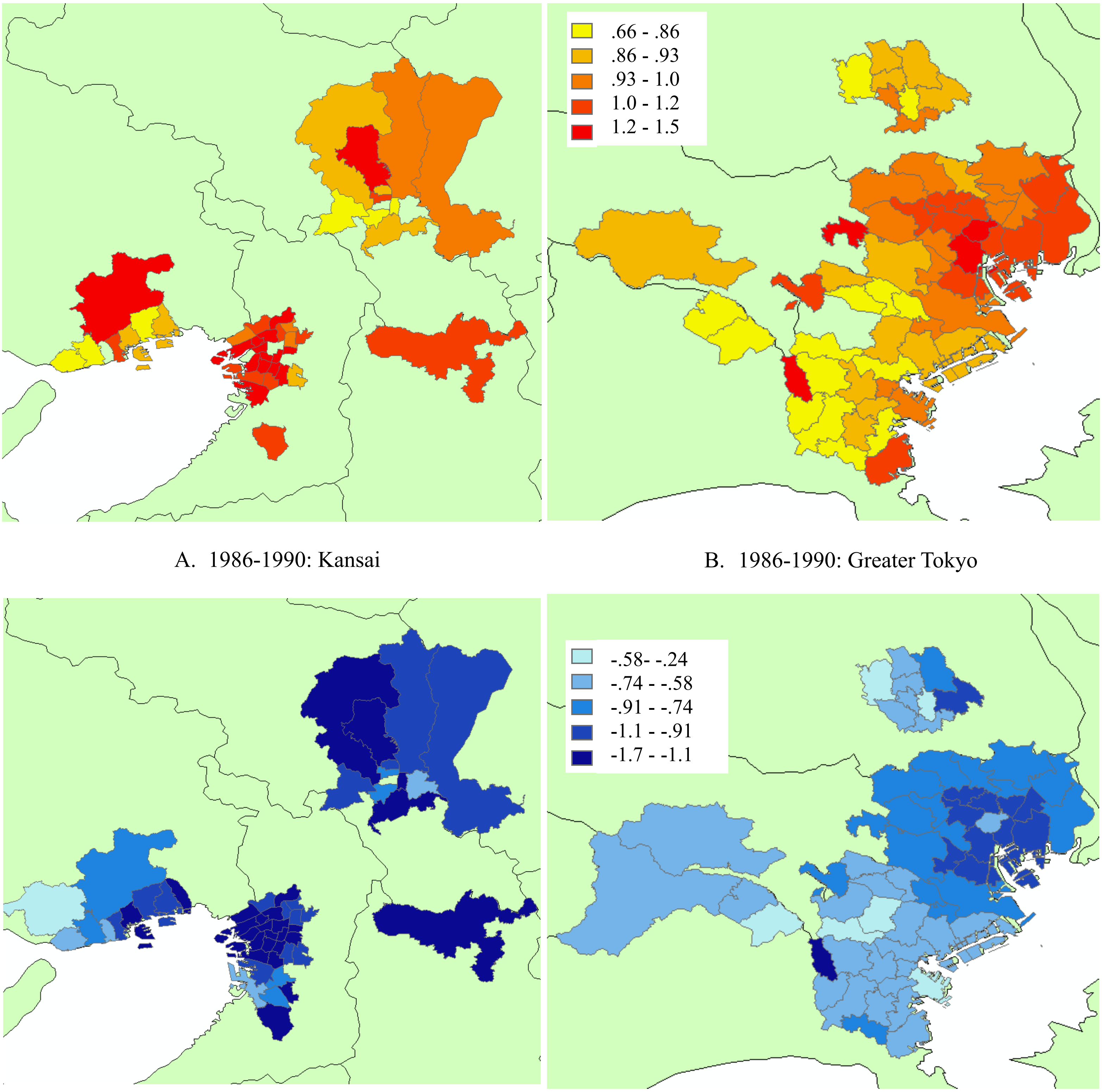


E. Time to CBD (30 min): Greater Tokyo



F. Time to CBD (30 min): Kansai

Figure 7. Estimated Counterfactual Density, by Year



A. 1986-1990: Kansai

B. 1986-1990: Greater Tokyo

C. 1990-2000: Kansai

D. 1990-2000: Greater Tokyo

Figure 8: Appreciation Rate of Log Price, by City