Summary

In this paper, we aim to analyze the relationships between the quality and price of secondhand condominiums in the 23 wards of Tokyo. We propose a secondhand condominium price estimation method. Specifically, with a linear model as the starting point, we estimate nonlinear relationships of secondhand condominium unit prices in terms of floor space in square meters, the number of years since construction, etc. A nonparametric estimation model using a continuous dummy variable (DmM) and a switching regression model (SWR) as a parametric estimation method were used on the basis of market structural changes. In addition, a generalized additive model (GAM) was applied for semiparametric estimations. These models were compared with the estimated linear model as the starting point. DmM, SWR, and GAM not only increased the power of explanation of all the models but also yielded similar nonlinearity estimates for the relationship between secondhand condominium quality and secondhand condominium unit price.

Keywords: Hedonic approach, nonlinearity, structure change, switching regression model, generalized additive models.

JEL Code: C31 - Cross-Sectional Models; Spatial, R31 - Housing Supply and Markets.
1. Objectives of the study

This study deals with an issue related to the “quality of property and information on the appropriate comparable price” in the Japanese secondhand condominium market. If an analysis is made, particularly of a large area or using a great deal of data, it should be assumed that there are suppliers and consumers who have various preferences in the market. However, many economic models often assume homogeneous market participants. Because of this, statistical models are assumed to be linear models in most cases. It is difficult to assume, however, that there are linear relationships between the main secondhand condominium pricing factors, including indices such as floor space in square meters, the number of years since construction, the time to the nearest station, and the travel time to the central business district, and secondhand condominium unit prices.

Regarding floor space, for instance, relatively small dwelling units will be purchased for investment purposes or by single households, slightly larger dwelling units will be purchased by small households such as the so-called double income no kids (DINK) households, and dwelling units larger than a certain size will be purchased by family households. It is difficult to assume that these different households have the same preferences and that the size of the dwelling unit and its unit price have a simple linear relationship (Asami and Ohtaki, 2000; Thorsnes and McMillen, 1998).

Regarding the number of years since construction, it is natural to expect that secondhand condominium unit prices will decrease with time after construction. This is because the physical deterioration of secondhand condominiums progresses with time, and in particular because deterioration associated with the recent marked advances in condominium technology affects the prices. It is further expected that bid prices will differ between consumers who prefer new

1
equipment and those who do not, or by income constraint, i.e., between high-income households and low-income households. Because a depreciation curve for the number of years since construction is a particularly important indicator of collateral assessment for housing loans, an earlier study focusing only on this variable was reported (Clapp and Giaccotto, 1998).

With regard to the time to the nearest station, areas near a railway station have more shops and more convenient transport links but often have an insufficient number of parks and a poor natural environment. Hence, households, including single and DINK households, expressing a preference for convenience will probably buy property in areas near a station, while family households, particularly those with children, will tend to select areas further from the nearest station. This factor is also considered to have the same effect as the floor space in square meters on secondhand condominium unit price.

The same holds true for the travel time to the central business district. Although it is assumed that secondhand condominium market prices diminish with distance from the central business district, households that have different preferences are expected to choose secondhand condominiums according to distance zones or regions. It is also difficult to assume that bid prices among these households with different preferences show continuous and linear changes.

The above problem is conventionally an issue of selecting the appropriate function form in the estimation of a hedonic function (Box and Cox, 1964; Cropper and McConnel, 1988; Halvorson and Pollakowski, 1981; Rasmussen and Zuehlke, 1990; Wooldridge, 1992; Pace, 1998).

In this study, we carry out hedonic estimation of the Tokyo metropolitan area secondhand condominium market focusing on four variables, namely, floor space in square meters, the number of years since construction, the time to the nearest station, and the travel time to the central business district, which are the main secondhand condominium pricing factors. With a linear model (OLS) as a starting point, a nonparametric estimation model using a continuous
dummy variable (DmM), a switching regression model (SWR) assuming change points of the
price structure, and a generalized additive model (GAM) which are capable of estimating the
nonlinearity of the secondhand condominium price structure. Clear linear relationships between
the four major variables and secondhand condominium prices were detected. DmM, SWR, and
GAM not only showed an improved power of explanation compared with the linear model but
also estimated similar nonlinearity of the secondhand condominium price structure in terms of the
variables.

In section 2, the OLS and the DmM, SWR, and GAM models, which take into account the
nonlinearity, are set up, and in section 3, the data used in this study are explained. In section 4,
estimations using the above-mentioned models are carried out, and in section 5, the results
estimated using the models are compared.

2. Causality between secondhand condominium price and quality

2.1. Hedonic approach: construction of a base model

In this section, a hedonic function for the Tokyo metropolitan area secondhand condominium
market is formulated, and the causality between price and quality in the secondhand
condominium market in the 23 wards of Tokyo is thereby analyzed.

The simplest model is set up below as the base model.

\[
\log RP / FS = a_0 + \sum_{h} a_{1h} \log X_{h} + \sum_{i} a_{2i} \log Z_{i} + \sum_{j} a_{3j} \cdot LD_{j} + \sum_{k} a_{4k} \cdot RD_{k} + \sum_{l} a_{5l} \cdot TD_{l} + \varepsilon
\]  

(1)

\[RP:\text{ Resale price of condominium (yen)}\]
\[X_{h}:\text{ Main variables}\]
\[FS:\text{ Floor space (square meters)}\]
This model for explaining secondhand condominium prices incorporates the floor space ($FS$), the number of years since construction ($Age$), the time to the nearest station ($TS$), and the travel time to the central business district ($TT$), which are variables considered to have important effects on secondhand condominium prices. In addition to these variables, information such as the amount of balcony space ($BS$) and the number of units ($NU$), available from magazines containing information on properties, was also incorporated in the models. Such information also relates to the condominium location or the building characteristics.

In addition to the above, “market reservation time,” that is, the time from the first appearance in a property magazine to the completion of the contract, was added. If the asking price is too high, or when secondhand condominiums are in oversupply compared with demand, leading to a glut in the market, or if a negative assessment is made of a secondhand condominium on the market on the basis of unpublished information not available in property magazines, the market reservation time is prolonged, and this effect is expected to be estimated as $\frac{\partial (RP/FS)}{\partial RT} > 0$. The market reservation time tends to become longer in the case of a thin market. In this case,
∂(RP/FS)/∂RT > 0 does not always hold, and variations in the conclusion price tend to be greater. This is an indicator determined not by the condominium location or building characteristics but by market conditions.

The railway line dummy (RDk) takes regional characteristics into account, and the time dummy (TDl) takes temporal changes in the market into account.

The objective of this study is to clarify the price structure of secondhand condominiums in the 23 wards of Tokyo, with the analysis focusing on the four main variables, FS, Age, TS, and TT.

2.2. Nonlinearity of housing prices

2.2.1. Estimation model of qualitative price structure disparities using dummy models

In the base model, it was assumed that the relationships between the main variables and secondhand condominium prices are linear. Experience shows, however, that it is difficult to assume that the main variables and unit prices have linear relationships (including a log-linear relationship), and that there is a possibility of kinks in the relationships at specific points. Specifically, the secondhand condominium price structure may vary depending on whether or not the location is within a bus service operating area when considering TS, and on the timing of amendments to the building code and large-scale repair when considering Age. We can also consider that there is an issue of different bid-price curves because of a mixture of different market participants, as mentioned in section 1. This issue belongs to a “structural-change problem” with a statistical model.

It is unknown, however, at what point the secondhand condominium price structure undergoes changes relative to each variable, and it is impossible to estimate prices by introducing structural-change points exogenously. As stochastic methods of dealing with such problems, those using
nonparametric models are available. For example, Bin(2004), Clapp(2003), Gencay and Yong (1996) and Thorsnes and McMillen (1998) used semiparametric methods to estimate such specific variables.

In the present study, the four main variables used as parametric variables in the base model were made into dummy variables to estimate the relationships between these variables and secondhand condominium unit prices. In forming the dummy variables, an arbitrary bandwidth ($\beta$) was set up for each variable unit. In setting the bandwidth ($\beta$), it is necessary to investigate the band unit with which consumers make a decision. The model obtained by forming the dummy variables based on the four main variables is referred to as the dummy model (hereafter referred to as DmM), and is formulated as shown below.

$$
\log \frac{RP}{FS} = a_0 + \sum \alpha_i \log Z_i + \sum \beta_j \cdot LD_j + \sum \gamma_k \cdot RD_k + \sum \delta_l \cdot TD_l + \sum \sigma \cdot Dm\left(FS_\sigma\right) \\
+ \sum a_{6\sigma} \cdot Dm\left(Age_\sigma\right) + \sum a_{7\gamma} \cdot Dm\left(TS_\gamma\right) + \sum a_{8\delta} \cdot Dm\left(TT_\delta\right) + \epsilon
$$

: $Dm(X_h)$: Dummy variables with bandwidth($\beta$) by main variables($X_h$)  \hspace{1cm} (2)

2.2.2. Development of a model by taking into account structural disparities – switching regression model

In the DmM, the main variables are set as dummy variables and are estimated on the basis of DmM being a nonparametric model. This model assumes that the secondhand condominium price structure changes continuously by the bandwidth ($\beta$) unit set for each variable. In the actual market, however, it is difficult to assume that the structure changes in all the bandwidth units. Also, in estimation, it is difficult to interpret estimated results because sequential changes in the
error structure are assumed. In other words, although DmM is applicable to any nonlinear-structure object, and although it is very good for observing the state of approximate price changes, it is necessary to analyze the obtained results by cross-checking errors for each dummy, because the errors of the estimated statistics of the continuously changing dummy variables are not uniform.

In this paper, the dummy model is simplified and it is assumed that there are two points of structural change and three price-bidding curves. For example, if attention is paid to the location-centered attribute, the following three types of purchasers are expected: i) one-room condominiums in which a single person lives, ii) compact-room condominiums in which a small household (for instance, a couple) lives and iii) relatively large condominiums in which a family lives. Depending on such differences in the entities, the single person and the small household will probably select a more convenient region, while the family will tend to attach more importance to the living environment. Hence, the $TS$ is also considered to be divided into the following: i) an area in which households attaching more importance to convenience are located, ii) an area that is within walking distance of the station but in which a good living environment is maintained and iii) an area that requires the use of buses or a car to reach the nearest station.

Given the above perspective, it may be thought that on the whole, three groups having three different preferences exist.

As described above, if the market structure is divided into three markets, two structural-change points should exist for the variable group thought to have nonlinearity. It is unknown, however, at what point the market structure changes (Jushan and Perron, 1998). Under such circumstances, the basic model is modified, and an estimate is made by exploratory analysis of variables thought to have nonlinearity, i.e., $FS$, $Age$, $TS$, and $TT$. Specifically, the following two dummy variables
are introduced on the assumption that the market is divided at points $l$ and $m$ for each main variable $X_h$. 

\[ D_{m(l_h\leq X_h < m_h)} : \text{if } l_h \leq X_h < m_h, \text{ then 1, otherwise 0} \]

\[ D_{m(m_h \leq X_h)} : \text{if } m_h \leq X_h, \text{ then 1, otherwise 0} \]

$l < m$

A model such as the one shown below is estimated by introducing the above dummy variables.

\[
\log \frac{RP}{FS} = a_0 + \sum_{l} a_{l_h} \log X_h + \sum_{j} a_{j_l} \log Z_j + \sum_{k} a_{k_m} \cdot LD_k + \sum_{s} a_{s_i} \cdot TD_s + \sum_{l} a_{s_l} \cdot TD_l
\]

\[ + \ a_6 \ \left( D_{m(l_h\leq X_h < m_h)} \right) + a_7 \ \left( D_{m(m_h \leq X_h)} \right) \ (\log X_h \ \left( D_{m(l_h\leq X_h < m_h)} \right)) \]

\[ + \ a_8 \ \left( D_{m(m_h \leq X_h)} \right) + \varepsilon \]

(3)

This model is the switching regression model (hereafter referred to as SWR), and we assume that the regression model is switched at points $l$ and $m$. For further information on SWR, see Shimizu and Nishimura (2007).

2.2.3. Generalized additive model

DmM estimates the nonlinear shape by an exploratory approach, and SWR, which assumes a price formation structure in which three price-bidding entities exist, is capable of estimating the nonlinearity. However, arbitrariness remains in DmM in the setting of the bandwidth ($\beta$), and the relationship between the secondhand condominium unit price and quality is treated as discontinuous. SWR is an effective method because it can extract differences in the price-bidding structure by an exploratory approach and approximate them as a nonlinear model if segments in
which bid prices are different are obvious. However, SWR needs to be based on strong assumptions regarding the number of price-bidding entities. If the current housing market of Japan is assumed, the assumption that the market is classified into three entities is considered to be reasonable. In the process of diversification of the social structure in the future, however, a more general model of SWR will be required to estimate the secondhand condominium price structure.

In this paper, to predict the relationship between the secondhand condominium unit price and quality mainly on the basis of data, we also estimate the secondhand condominium price structure using a generalized additive model (Hastie and Tibshirani, 1986, 1990), hereafter referred to as GAM.

A GAM is a generalized linear model (Nelder and Wedderburn, 1972) with a linear predictor involving a sum of smooth functions of the covariates. A generalized linear model has the basic structure:

$$g(\mu) = \beta'x,$$  \hspace{1cm} (4)

where $\mu \equiv E(Y)$ and $Y$ is independent ($Y \sim$ some exponential family distribution), $g$ is a smooth monotonic link function, $x$ is the explanatory variable matrix, and $\beta$ is a vector of unknown parameters. In general, GAMs have a structure with smooth function as follows:

$$g(\mu) = \beta'x + s_1(X_1) + s_2(X_2) + \cdots s_H(X_H),$$  \hspace{1cm} (5)

where $s_h$ are smooth functions of the covariates $X_1, X_2, \cdots, X_H$. The model allows for rather flexible specification of the dependence of response on covariates. Smooth functions are represented using penalized splines with smoothing parameters selected by generalized cross-
validation (GCV). Graphs of the estimated smooth function versus $X_h$ constitute one of the main products of the GAM estimator.

The estimation model of hedonic function with GAMs is shown below.

$$\log \frac{RP}{FS} = a_0 + \sum_h s_h(X_h) + \sum_i a_{2i} \log Z_i + \sum_j a_{3j} \cdot LD_j + \sum_k a_{4k} \cdot RD_k + \sum_l a_{5l} \cdot TD_l + \epsilon$$  \hspace{1cm} (6)

where $\log \frac{RP}{FS}$ is the identity link, and $s_h(X_h)$ is an unspecified smooth function for the main variable $(FS, Age, TS, TT)$. The other terms are predictors ($\log Z, LD, RD, TD$) via a linear combination term with parameters $a_2, a_3, a_4, a_5$.

3. Data

3.1. Characteristics of secondhand condominium price data

This study used information published in the “Shukan Jutaku Joho” (Weekly Housing Information Magazine) by Recruit Co., Ltd. This magazine provides weekly information on housing quality and asking prices, and it includes historical information from the first appearance on the market via the publication up to deletion of the information because of the completion of the contract or for other reasons. The information consists roughly of information on the asking price of a secondhand condominium at the time of appearance on the market, information on the price of secondhand condominiums at the time of deletion from the magazine, and information on transaction prices collected as samples.

It has been pointed out that it is generally better to use transaction price information to analyze the real-estate market. However, this is a comparison with assessment information such as the posted price. Because real-estate transactions are ultimately relative transactions, the individual
circumstances of sellers (rush selling) and buyers (rush buying) are often involved in transaction prices. Because of this, it can hardly be said that in current housing market transactions, which contain inefficiency, that the quality and marketability of transaction price information are high. When transaction price information is used in actual real-estate assessment, data compensated for the circumstances are used, but it is difficult to understand quantitatively the circumstances surrounding real-estate deals. It cannot be said that the marketability of this data is higher than the real-estate assessment information.

In this study, it was decided to use the information at the time of deletion from Weekly Housing Information, i.e., the price at the completion of the contract. From the time of putting out information on the quality and price in the magazine, the asking price of a property is reduced, i.e., the reverse-auction approach, until buyers appear; therefore, the prices at the time of deletion from the magazine have the characteristic of being the upper limit of buyers’ bid prices because they are the first accepted prices. However, these prices are relative prices formed in a competitive market compared with information on real-estate deal prices, and, because of this, the prices can be said to be a suitable indicator that does not include individual circumstances surrounding the deals but reflects the market environment and consumers’ individual preferences. Thus, we conducted analyses using the condominium price information at the time when the information was deleted from the magazine.

3.2. Preparation of analytical data

In this study, we used secondhand condominium data, the listing of which was deleted from the magazine between the first week and the last week of 2005. A list of the data analyzed is shown in Table 1.
### Table 1. List of analyzed data.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Variables</th>
<th>Contents</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>Floor space/ square meters</td>
<td>Floor space.</td>
<td>m²</td>
</tr>
<tr>
<td>AGE</td>
<td>Age of Building: Number of years since construction</td>
<td>Period between the date when the data is deleted from the magazine and the date of construction of the building.</td>
<td>year</td>
</tr>
<tr>
<td>TS</td>
<td>Time to nearest station</td>
<td>Time distance to the nearest station (walking time), Period between the date when the data is deleted from the magazine and the date of construction of the building.</td>
<td>minute</td>
</tr>
<tr>
<td>TT</td>
<td>Travel Time to central business district</td>
<td>Minimum of railway riding time in daytime to Terminal 7 stations in 2005*.</td>
<td>minute</td>
</tr>
<tr>
<td>BS</td>
<td>Balcony space/ square meters</td>
<td>Balcony space.</td>
<td>m²</td>
</tr>
<tr>
<td>NU</td>
<td>Number of units</td>
<td>Total units of the condominium.</td>
<td>unit</td>
</tr>
<tr>
<td>RT</td>
<td>Market reservation time</td>
<td>Period between the date when the data appear in the magazine for the first time and the date of being deleted.</td>
<td>week</td>
</tr>
<tr>
<td>FF</td>
<td>First floor dummy</td>
<td>The property is on the ground floor 1, on other floors 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>HF</td>
<td>Highest floor dummy</td>
<td>The property is on the top floor 1, on the other floors 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>SD</td>
<td>South-facing dummy</td>
<td>Fenestrae facing south 1, other directions 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>FD</td>
<td>Ferroconcrete dummy</td>
<td>Steel reinforced concrete frame structure 1, other structure 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>LDj</td>
<td>Location (Ward) dummy</td>
<td>jth administrative district 1, other district 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>RDk</td>
<td>Railway line dummy</td>
<td>kth railway line 1, other railway line 0.</td>
<td>(0,1)</td>
</tr>
<tr>
<td>TDi</td>
<td>Time dummy (monthly)</td>
<td>lth month 1, other month 0.</td>
<td>(0,1)</td>
</tr>
</tbody>
</table>

*Terminal Staion : Tokyo, Shinagawa, Shibuya, Shinjuku, Ikebukuro, Ueno, and Ootemachi

Transportation convenience at each point was first represented by TS and TT. Regarding TS, only data related to condominiums within walking distance were extracted, and the walking time to the nearest station (in minutes) was adopted.
Regarding the TT, the mean daytime travel time from the nearest station to the main terminal stations on the Yamanote line (Tokyo Loop Line), specifically Tokyo, Shinagawa, Shibuya, Ikebukuro, Shinjuku, and Ueno, as well as Otemachi (a major subway station), were measured, and the minimum time was adopted. By statically incorporating the marketability of each object, we can obtain its price, but if a time element is incorporated, the period between the date of the first appearance of data in the magazine and the date of completing the contract is also included.

Data were prepared by considering this time—from the first appearance in the magazine to deletion from it upon the completion of the contract—as the market reservation time (RT).

As information regarding the quality of secondhand condominiums themselves, information on FS, Age, BS, and NU were used. In addition, regarding items considered to affect condominium value, including whether condominiums are on the highest floor, whether they are on the ground floor, and the direction in which windows face (assuming that the prices of condominiums with windows facing south are higher, a “south-facing dummy” was set), data for analysis were obtained using relevant data published in the housing magazine, and dummy variables were produced from the information.

3.3. Statistical characteristics of analytical data

Table 2 shows descriptive statistics of the main variables.

The average secondhand condominium price was 32.53 million yen, the minimum price was 8.5 million yen, and the maximum price was 188 million yen, with a standard deviation of 18.58 million yen, showing the considerable variation in the price of condominiums. Condominiums from small one-room units to luxurious, expensive units were included. The average unit price
per m² was approximately 510,000 yen/m², and a right-skewed but unimodal distribution of the prices is observed (Fig. 10).

Table 2. Summary of statistical values of secondhand condominium price data.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP: Resale Price of Condominium (10,000 Yen)</td>
<td>3,253.89</td>
<td>1,858.83</td>
<td>850.00</td>
<td>18,800.00</td>
</tr>
<tr>
<td>FS: Floor space (㎡)</td>
<td>61.82</td>
<td>19.83</td>
<td>16.00</td>
<td>133.29</td>
</tr>
<tr>
<td>RP/FS</td>
<td>51.71</td>
<td>17.93</td>
<td>24.01</td>
<td>164.08</td>
</tr>
<tr>
<td>Age: Age of Building (year)</td>
<td>16.51</td>
<td>9.92</td>
<td>0.42</td>
<td>34.42</td>
</tr>
<tr>
<td>TS: Time to the nearest station: Walk Time (minutes)</td>
<td>7.45</td>
<td>4.19</td>
<td>1.00</td>
<td>26.00</td>
</tr>
<tr>
<td>TT: Travel Time to Central Business District (minutes)</td>
<td>14.83</td>
<td>5.23</td>
<td>0.00</td>
<td>30.00</td>
</tr>
<tr>
<td>BS: Balcony Space (㎡)</td>
<td>8.14</td>
<td>5.96</td>
<td>0.00</td>
<td>80.94</td>
</tr>
<tr>
<td>NU: The Number of Units</td>
<td>88.03</td>
<td>122.48</td>
<td>10.00</td>
<td>1149.00</td>
</tr>
<tr>
<td>RT: Market reservation time (week)</td>
<td>9.33</td>
<td>8.37</td>
<td>1.00</td>
<td>64.00</td>
</tr>
</tbody>
</table>

For TS, only the distribution with respect to time was examined in this paper, and there were many condominiums in very convenient locations. The minimum time was 1 minute, the maximum time was 26 minutes, and the average time was 7.45 minutes. This indicated that condominiums were constructed with importance placed on the convenience of their location. The average TT is 14 minutes and the maximum TT is 30 minutes. TT is the minimum value of average travel times during the daytime from the nearest station to seven main terminal stations.
The distribution of TT (Fig. 10) is right skewed with samples for which TT is 0 minutes and the next shortest TT is 10 minutes. In other words, this distribution shows one group of condominiums located at main terminal stations where TT is 0 minutes, and another group for which the minimum TT is 10 minutes. Attention should be paid to this unique distribution of TT.

Regarding FS, the minimum FS was 16.00 m² and the maximum FS was 133.29 m², with an average FS of 61.82 m², and condominiums ranging from those for single-person households to large condominiums are included.

Regarding Age, the average age was 16.3 years, and the maximum age was 39.67 years, showing that the secondhand condominiums were mainly old ones. When attention is paid to their distribution, there was more than one peak, including one within the new condominium bracket of less than 5 years since construction and another peak within the older condominium bracket. It is expected that the secondhand condominiums built during the latter half of the 1980s to the early 1990s have a lower rate of appearance in the market or of buyers.

4. Estimation results

4.1. Estimation of base model

First, a base model is estimated as the simplest model in terms of unit resale price per m², as shown below. This model is estimated to have a relatively high power of explanation with a determination coefficient adjusted for the degrees of freedom of 0.775 (see Table 3 for details).
\[
\log \frac{RP}{FS} = 3.931 + 0.047 \cdot \log FS - 0.188 \cdot \log Age - 0.054 \cdot \log TS - 0.117 \cdot \log TT + 0.012 \cdot \log BS \\
(155.26) \quad (+8.98) \quad (−96.38) \quad (−21.51) \quad (−5.24) \quad (12.13)
\]

\[
+ 0.020 \cdot \log NU - 0.006 \cdot RT - 0.034 \cdot FF + 0.054 \cdot HF - 0.012FD + 0.003SD \\
(15.33) \quad (−3.33) \quad (−6.20) \quad (5.37) \quad (−3.23) \quad (0.965)
\]

\[
+ \beta_{1h} \sum LD_h + \beta_{2i} \sum RD_i + \beta_{3j} \sum TD_j
\]

Adjusted R-Square: 0.775 (t-values shown in parentheses)

Number of observations: 9,682

Because this base model was based on data pooled for a year, point modification was accomplished by the forced input of a time dummy (TD), and the secondhand condominium price structure was estimated using the property characteristics of secondhand condominiums and the railway/subway dummy (RD). Among the property characteristics, FS, BS and NU were estimated positively, while Age, TS and TT were estimated negatively.

First, regarding FS, the construction cost per square meter is expected to diminish with an increase in condominium size, but if condominium size and grade are positively correlated, the unit resale price increases with increasing size. This holds true for BS, because it is expected that BS tends to become larger with increasing condominium grade. Because of this, BS is considered to be positively estimated for the condominium scale. The NU is a representative index showing the resale price of a condominium as a whole rather than the price of each unit price. For example, because shared space tends to be ampler with increasing NU, this space is considered to affect the condominium unit price.

With the recent rapid progress in condominium technology, the functionality of secondhand condominiums declines, and large-scale repair costs increase, with time. Hence, the condominium resale price decreases as Age increases. In addition, convenience in commuting to
offices or schools decreases with increasing \( TS \). Because there are fewer shops and services and daily life is less convenient far from stations, condominium resale prices are expected to decrease. Furthermore, because, in general, more people commute to the central business district, not only commuting expenses but also commuters’ opportunity costs increase, which is thought to contribute to decreased condominium prices.

Because there are not only the above-mentioned real-estate-specific factors but also broad disparities in the housing environment among administrative municipalities or areas along railway lines, which cannot be considered in the functions estimated in this paper, such disparities are estimated using the dummy variables.

With the base model as the starting point, it is modified to give the other estimation models below. In concrete terms, regarding the variables, except for those to be improved as instrumental variables, all those adopted for the base model are forcibly incorporated in the other models.
Table 3. Estimation results using the base model

Method of Estimation
OLS

Dependent Variable
RP: Resale Price of Condominiums (in log)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property Characteristics (in log)</td>
<td></td>
<td></td>
<td>Railway/Subway Line Dummy</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>3.931</td>
<td>155.275</td>
<td>RDj (j=0,…,J)</td>
<td>0.033</td>
</tr>
<tr>
<td>FS: Floor space</td>
<td>0.047</td>
<td>8.984</td>
<td>Yamanote</td>
<td>0.056</td>
</tr>
<tr>
<td>Age: Age of building</td>
<td>-0.188</td>
<td>-96.379</td>
<td>Ginza</td>
<td>0.158</td>
</tr>
<tr>
<td>TS: Walk Time to the nearest station</td>
<td>-0.054</td>
<td>-21.510</td>
<td>Marunouchi</td>
<td>0.085</td>
</tr>
<tr>
<td>TT: Travel Time to CBD</td>
<td>-0.017</td>
<td>-5.257</td>
<td>Hibiya</td>
<td>0.040</td>
</tr>
<tr>
<td>BS: Balcony space</td>
<td>0.012</td>
<td>4.471</td>
<td>Tozai</td>
<td>0.053</td>
</tr>
<tr>
<td>NU: Number of units</td>
<td>0.020</td>
<td>10.190</td>
<td>Yurakucho</td>
<td>0.067</td>
</tr>
<tr>
<td>RT: Market reservation time</td>
<td>-0.006</td>
<td>-3.331</td>
<td>Yotsuka</td>
<td>0.053</td>
</tr>
<tr>
<td>Property Characteristics (dummy variables)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF: First Floor Dummy</td>
<td>-0.034</td>
<td>-6.198</td>
<td>Oedo</td>
<td>-0.026</td>
</tr>
<tr>
<td>HF: Highest floor dummy</td>
<td>0.054</td>
<td>5.365</td>
<td>TokyoMonorail</td>
<td>-0.338</td>
</tr>
<tr>
<td>FD: Ferroconcrete dummy</td>
<td>0.012</td>
<td>-5.226</td>
<td>Keikyu</td>
<td>-0.214</td>
</tr>
<tr>
<td>SD: South-facing dummy</td>
<td>0.003</td>
<td>0.965</td>
<td>Keikyuokubo</td>
<td>-0.265</td>
</tr>
<tr>
<td>Location (Ward) Dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDi (i=0,…,I)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiyoda</td>
<td>0.550</td>
<td>33.970</td>
<td>Yokosuka</td>
<td>-0.089</td>
</tr>
<tr>
<td>Chuo</td>
<td>0.257</td>
<td>21.513</td>
<td>TokyoAki</td>
<td>-0.025</td>
</tr>
<tr>
<td>Minato</td>
<td>0.602</td>
<td>44.850</td>
<td>TokyoYoyoku</td>
<td>0.076</td>
</tr>
<tr>
<td>Shinjuku</td>
<td>0.384</td>
<td>37.845</td>
<td>Denenchoji</td>
<td>0.032</td>
</tr>
<tr>
<td>Bunkyo</td>
<td>0.323</td>
<td>30.172</td>
<td>Yokosuka</td>
<td>-0.045</td>
</tr>
<tr>
<td>Shibuya</td>
<td>0.031</td>
<td>2.178</td>
<td>Odaikyo</td>
<td>0.015</td>
</tr>
<tr>
<td>Sumida</td>
<td>0.003</td>
<td>0.965</td>
<td>Keieikubojou</td>
<td>0.004</td>
</tr>
<tr>
<td>Shinagawa</td>
<td>0.396</td>
<td>33.970</td>
<td>Keiseigo</td>
<td>-0.126</td>
</tr>
<tr>
<td>Meguro</td>
<td>0.499</td>
<td>36.090</td>
<td>Cyuuou</td>
<td>0.065</td>
</tr>
<tr>
<td>Ota</td>
<td>0.266</td>
<td>22.863</td>
<td>Seibuen</td>
<td>-0.063</td>
</tr>
<tr>
<td>Setagaya</td>
<td>0.430</td>
<td>32.067</td>
<td>Saikyo</td>
<td>-0.073</td>
</tr>
<tr>
<td>Shibuya</td>
<td>0.619</td>
<td>57.615</td>
<td>Tobulenesaki</td>
<td>-0.111</td>
</tr>
<tr>
<td>Nakano</td>
<td>0.256</td>
<td>21.462</td>
<td>KeiseiOshiu</td>
<td>-0.125</td>
</tr>
<tr>
<td>Suganami</td>
<td>0.215</td>
<td>22.522</td>
<td>Yukarigoko</td>
<td>0.056</td>
</tr>
<tr>
<td>Toshima</td>
<td>0.234</td>
<td>18.304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hira</td>
<td>0.089</td>
<td>5.553</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arakawa</td>
<td>0.080</td>
<td>4.374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nerima</td>
<td>0.101</td>
<td>9.454</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adachi</td>
<td>-0.027</td>
<td>-9.546</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Katsushika</td>
<td>-0.110</td>
<td>-6.491</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edogawa</td>
<td>-0.067</td>
<td>-4.914</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adjusted R square= 0.775
Number of Observations= 9,682
4.2. Estimation using continuous-quality dummy models (DmM)

Model estimation was conducted using continuous dummy variables in accordance with the model shown in equation 2. Here, in forming a dummy variable corresponding to each main variable, the problem was how to set its bandwidth ($\beta$). For example, regarding $FS$, it is unlikely that consumers change their preference based on 1 m$^2$ as a unit, and because of this, the bandwidth was set as $\beta = 5$. It was considered unproblematic to set $\beta = 1$ for the $Age, TS$ and $TT$.

$$Dm(FS_{\rho}) : \rho = 15,20,25,30,\ldots,135$$

$$Dm(Age_{\sigma}) : \sigma = 1,2,3,4,5,\ldots,35$$

$$Dm(TS_{\zeta}) : \zeta = 1,2,3,4,5,\ldots,30$$

$$Dm(TT_{\tau}) : \tau = 1,2,3,4,5,\ldots,30$$

The estimation results are shown below.

$$\log \frac{RP}{FS} = 3.962 + 0.007 \cdot \log Age + 0.032 \cdot \log NU - 0.007 \cdot RT - 0.044 \cdot FF + 0.055 \cdot HF$$

$$- 0.015 \cdot FD + 0.007 \cdot SD + \beta_{h} \sum h LD_{h} + \beta_{j} \sum j RD_{j} + \beta_{j} \sum j TD_{j} + \beta_{\rho} \sum \rho Dm(FS_{\rho})$$

$$+ \beta_{\sigma} \sum \sigma Dm(Age_{\sigma}) + \beta_{\zeta} \sum \zeta Dm(TS_{\zeta}) + \beta_{\tau} \sum \tau Dm(TT_{\tau})$$

Adjusted R-Square: 0.817 (t-value shown in parentheses)

Number of observations: 9,682

The determination coefficient adjusted for the degrees of freedom was improved greatly from 0.775 for the base model to 0.819. There was no great change in the estimated statistics for the
other parametric variables except for the code relations of the dummy variables and the continuous dummies, compared with those in the base model.

The estimation results for the dummy variables are shown in Table 7.

4.3. Expansion to a model coupled with structural disparity – switching regression model

4.3.1. Estimation of individual index models

In the function that was estimated as the base model, it was assumed that there was a simple linear relationship between the unit resale price and each variable. However, in actuality, it was difficult to assume that each variable had a simple linear relationship with the unit resale price.

An estimation was made, therefore, of the relationships of the four variables having the greatest impact on secondhand condominium resale price formation, i.e., Age, TS, FS, and TT, with the unit resale price using the dummy variables set in the previous section. On the basis of estimation results, it was understood that, for Age, TS, and FS, there were specific kinks in the relationship between unit resale price and each factor and that, for the FS, sign reversal occurred in some cases. Because the areas of analysis were city-center districts with an extensive transportation system and variations in travel time were not very large, TT seemingly had a greater impact on the unit resale price with increasing TT, but the unit resale price was found to decrease linearly with time.

From the above results, assuming that points of price structure change exist in the relationship between unit resale price and condominium quality, structural estimation was carried out using the SWR. Here, on the assumption that there were two such points in each relationship, these change points were explored. Under ordinary circumstances, l and m were set for every main variable $X_h$. Because it was difficult to optimize them simultaneously, optimization was carried
out for each variable using the base model as a starting point. A model assessment was performed on the basis of Akaike’s information criterion (AIC).

To confirm whether a structural change occurred in the detected $l$ and $m$ following the above model estimation, a structural-change test was conducted using the F-test.

**Estimation results for $FS$**

$FS$ was changed by units of 5 m$^2$ for consistency with DmM. The ranges of combinations of $l$ and $m$ in $Dm_{(l\leq Xh< \text{mth})}$ and $Dm_{(m<Xh)}$ were $l > 15$, $m < 135$, and $l < m$, and there were 253 combinations. By estimating all the 253 combinations of 253 functions, their AIC values were compared. Estimation results showed that AIC was minimum at $l = 40$ and $m = 90$, and the determination coefficient adjusted for the degrees of freedom was 0.779, showing an improvement in the power of explanation. Figure 1 shows the combinations of $l$ and $m$ and changes in AIC.

**Estimation results for $Age$**

On the basis of the distribution of data on $Age$, the range of analysis was from more than 1 year to 35 years. The range of combinations of $l$ and $m$ in $Dm_{(l\leq Xh< \text{mth})}$ and $Dm_{(m<Xh)}$ were $l > 2$, $m < 35$, and $l < m$, and there were 561 combinations. By estimating all the 561 combinations of 561 functions, their AIC values were compared. Estimation results showed that AIC was a minimum at $l = 12$ and $m = 23$, and the determination coefficient adjusted for the degrees of freedom was 0.801, showing an improvement in the power of explanation compared with that of the base model. Figure 1 shows the combinations of $l$ and $m$ and changes in AIC.
Estimation results for TS

On the basis of the distribution of data on TS, the range of analysis was from more than 1 minute to 30 minutes. The ranges of combinations of $l$ and $m$ in $D_{m(h \leq X < mh)}$ and $D_{(mh \leq X < h)}$ were $l > 2$, $m < 30$, and $l < m$, and there were 300 combinations. By estimating all the 300 combinations of 300 functions, their AIC values were compared. Estimation results showed that AIC was a minimum at $l = 12$ and $m = 17$, and the determination coefficient adjusted for the degrees of freedom was 0.777, showing an improvement in the power of explanation compared with that of the base model. Figure 1 shows the combinations of $l$ and $m$ and changes in AIC.

Estimation results for TT

On the basis of the distribution of data on TT, the range of analysis was from more than 0 minutes to 30 minutes. The ranges of combinations of $l$ and $m$ in $D_{m(h \leq X < mh)}$ and $D_{(mh \leq X < h)}$ were $l \geq 1$, $m < 30$, and $l < m$, and there were 406 combinations. By considering all the 406 combinations of 406 functions, their AIC values were compared. Estimation results showed that AIC was a minimum at $l = 11$ and $m = 15$, and the determination coefficient adjusted for the degrees of freedom was 0.777, showing an improvement in the power of explanation compared with that of the base model. Figure 1 shows combinations of $l$ and $m$ and changes in AIC.
4.3.2. Confirmation by structural-change test

Optimum models were selected from the possible combinations in the above estimation. However, there was no evidence that structural change occurred in the sections extracted here. To demonstrate the presence of this change, a structural-change test (an F-test) was conducted. Specifically, the three groups divided by \( l \) and \( m \) in \( Dm_{(l<h<\text{middle})} \) and \( Dm_{(m\leq h)} \) were subjected to an F-test. Group I was set as \( X_{(h<l)} \), group II was set as \( X_{(l<h<\text{middle})} \), and group III was set as \( X_{(m<h)} \).

Three tests were conducted, between group I and group II, group II and group III, and group I and group III, for each variable (\( FS, Age, TS \), and \( TT \)). It is particularly important to verify whether or not there was a structural change between group I and group II, and group II and group III. If a structural change can be verified, then there is a nonlinear relationship between the unit resale...
price and each variable. If the F-test detects structural change between group I and group II, and between group II and group III but not between group I and group III, then the structure is different only within $l < h < m$.

The results of the structural-change test showed that a structural change occurred at the two previously determined values of $l$ and $m$ for $FS$, $Age$, and $TS$ with a significant difference of 10%. For the $TT$, no structural change was observed between group I and group II, but a structural change was found to exist between group II and group III. A structural change was also observed between group I and group III. These findings indicated that the structure changed only for $m = 15$ minutes or more.

### Table 4. Test results for structural-change test (Prob>0)

<table>
<thead>
<tr>
<th>Variable</th>
<th>I vs. II $X_{(k&lt;l)}$ vs. $X_{(j&lt;k&lt;l)}$</th>
<th>II vs. III $X_{(k&lt;l)}$ vs. $X_{(j&lt;k&lt;l)}$</th>
<th>I vs. III $X_{(k&lt;l)}$ vs. $X_{(n&gt;h)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$FS$: Floor Space (㎡)</td>
<td>0.00003</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>$Age$: Age of Building(months)</td>
<td>0.00179</td>
<td>0.08101</td>
<td>0.05582</td>
</tr>
<tr>
<td>$TS$: Time to nearest Station(minutes)</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.01115</td>
</tr>
<tr>
<td>$TT$: Travel Time to Central Business District (minutes)</td>
<td>0.22236</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>
4.3.3. Model estimation coupled with structural change

In the individual index models (FS, Age, TS, and TT), two structurally different sections were estimated by an exploratory approach for FS, Age, TS, and TT. By extracting optimum structural-change sections for measurement using AIC as an assessment index, a structural-change test was conducted using the F-test. Two structural-change points were detected for FS, Age, and TS. It was found, however, that there was only one structural-change point for TT.

If FS, Age, TS, and TT, for which the structural-change points were explored, are independent, the optimization of the structurally changed sections is maintained even when the hedonic function is estimated by simultaneously inputting the independently determined \( l \) and \( m \). The model was formulated on the basis of equation 3 by coupling with the structurally changed sections extracted using the individual models. The estimation results using the model are shown below. Table 5 shows estimated statistics as constant-term dummies and estimated statistics obtained as cross-terms.

\[
\log \frac{RP}{FS} = 4.242 - 0.094 \cdot \log FS - 0.086 \cdot \log Age - 0.046 \cdot \log TS - 0.009 \cdot \log TT + 0.086 \cdot \log BS \\
(63.25) \quad (-4.89) \quad (-24.87) \quad (-16.93) \quad (-2.78) \quad (3.64)
\]

\[
+ 0.031 \cdot \log NU - 0.007 \cdot RT - 0.042 \cdot FF + 0.052 \cdot HF - 0.015FD + 0.007SD + \beta_1 \sum LD_h \\
(16.89) \quad (-4.44) \quad (-8.28) \quad (5.58) \quad (-4.50) \quad (2.28)
\]

\[
+ \beta_2 \sum RD_i + \beta_3 \sum TD_j + \beta_5 DM_{(l \leq X_l < mX_l)} + \beta_6 DM_{(mX_l \leq X_l < hX_l)} + \beta_7 (\log X_k) \bigg( DM_{(l \leq X_l < mX_l)} \bigg)
\]

\[
+ \beta_8 (\log X_k) \bigg( DM_{(mX_l \leq X_l < hX_l)} \bigg)
\]

Adjusted R-Square: 0. 812 (t-values shown in parentheses)

Number of observations: 9,682
Table 5. Estimation parameters related to structural disparities

**Dummy Effect**

<table>
<thead>
<tr>
<th>Dummy Effect</th>
<th>(Dm_{(l&lt;0,m&gt;0)})</th>
<th>Coefficient</th>
<th>t-value</th>
<th>(Dm_{(m&lt;0,l&gt;0)})</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dm:FS\ (l=40, m=90))</td>
<td>-0.387</td>
<td>-5.149</td>
<td>-1.374</td>
<td>-6.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:Age\ (l=12, m=23))</td>
<td>0.579</td>
<td>11.733</td>
<td>0.109</td>
<td>1.521</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:TS\ (l=12, m=17))</td>
<td>0.216</td>
<td>2.130</td>
<td>0.773</td>
<td>2.682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:TT\ (l=12, m=17))</td>
<td>—</td>
<td>—</td>
<td>0.458</td>
<td>10.901</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cross Term Effect**

<table>
<thead>
<tr>
<th>Cross Term Effect</th>
<th>(Dm_{(l&lt;0,m&gt;0)})</th>
<th>Coefficient</th>
<th>t-value</th>
<th>(Dm_{(m&lt;0,l&gt;0)})</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dm:FS\ (l=40, m=90))</td>
<td>0.110</td>
<td>5.188</td>
<td>0.339</td>
<td>6.711</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:Age\ (l=12, m=23))</td>
<td>-0.241</td>
<td>-14.059</td>
<td>-0.106</td>
<td>-4.831</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:TS\ (l=12, m=17))</td>
<td>-0.099</td>
<td>-2.522</td>
<td>-0.296</td>
<td>-2.993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dm:TT\ (l=12, m=17))</td>
<td>—</td>
<td>—</td>
<td>-0.163</td>
<td>-11.236</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the formulated model, the determination coefficient adjusted for the degrees of freedom was 0.812, a value that is a great improvement over the 0.775 for the base model. The estimated model has the power of explanation equivalent to that of DmM. The estimation parameters related to structural disparities are, in general, accurately estimated.

4.3.4. Estimation by generalized additive model

Next, a GAM estimation was made. The following model was set up as a semiparametric regression model containing both parametric terms and nonparametric terms for smoothing.

\[
\log(RP/FS) = a_0 + s_{FS}(FS) + s_{AGE}(AGE) + s_{TS}(TS) + s_{TT}(TT) + \beta'x + \epsilon \tag{6}
\]
$s_{FS}(FS)$, $s_{AGE}(AGE)$, $s_{TS}(TS)$ and $s_{TT}(TT)$ were modeled nonparametrically and were subject to smoothing. The $\beta'x$ term is parametric.

For comparison, the ordinary linear regression model below was set up.

$$\log(\frac{RP}{FS}) = \alpha + \gamma_{FS}FS + \gamma_{AGE}AGE + \gamma_{TS}TS + \gamma_{TT}TT + \beta'x + u$$

(7)

This served as the base model formulated in advance. R language (ver. 2.31) and package library `gam` and `mgcv` were used for GAM estimation.

The estimation results are as follows.

$$\log RP/FS = 3.475 + 0.008 \cdot \log BS + 0.032 \cdot \log NU - 0.007 \cdot RT - 0.043 \cdot FF$$

$$+ 0.054 \cdot HF - 0.016 \cdot FD + 0.008 \cdot SD + \hat{\beta}_1 \sum \frac{LD}{n} + \hat{\beta}_2 \sum RD_j$$

$$+ \hat{\beta}_3 \sum TD_j + s_{FS}(FS) + s_{AGE}(AGE) + s_{TS}(TS) + s_{TT}(TT)$$

Adjusted R-Square: 0. 814 (t-value shown in parentheses)
Number of samples: 9,682

In this model, the determination coefficient adjusted for the degrees of freedom was 0.814, a value that indicates that the estimated model has a power of explanation similar to those of DmM and SWR. Note that the powers of explanation of DmM, SWR, and GAM were improved equivalently compared with that of the base model.

Table 6 shows the estimated performance of the smoothing function by GAM. No great difference was observed in the coefficients other than the smoothed parameters in comparison with the base model (Table 9). The degrees of freedom of the smoothing term obtained in terms
of the GCV standard were not integral. The F-value is a statistic that shows whether there was any difference between the effect of smoothed cases and that of nonsmoothed cases, and it was indicated that smoothing produced a significant difference between the models.

Table 6. Estimated smoothing parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated d.f</th>
<th>F-statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{FS}(FS)$</td>
<td>7.573</td>
<td>33.41</td>
<td>0.000</td>
</tr>
<tr>
<td>$s_{Age}(Age)$</td>
<td>8.518</td>
<td>1366.12</td>
<td>0.000</td>
</tr>
<tr>
<td>$s_{TS}(TS)$</td>
<td>7.779</td>
<td>96.77</td>
<td>0.000</td>
</tr>
<tr>
<td>$s_{TT}(TT)$</td>
<td>8.983</td>
<td>26.72</td>
<td>0.000</td>
</tr>
</tbody>
</table>

GCV score: 0.019
Deviance explained: 81.60%
Number of Observations: 9,682

(Note: GCV score is an indicator of the error of the generalized cross-validation method, and the ‘deviance explained’ is an index of the applicability of the theoretical figure to the actual performance.)

5. Relationships between secondhand condominium prices and floor space ($FS$), number of years since construction ($Age$), time to nearest station ($TS$), and travel time to central business district ($TT$)

A comparison of the estimated shapes of curves for the variables based on the above-described results for a series of estimations revealed the following.

5.1. Relationship between floor space ($FS$) and unit prices

There is a strong nonlinear relationship between $FS$ and unit resale prices. In small condominiums with $FS$ of approximately 20 m$^2$, the marginal effect on the unit resale price per
m² area is high but gradually decreases. It was found, however, that when $FS$ is larger than 80 m², the marginal effect increases rapidly. This tendency is shown by DmM, SWR, and GAM but not by the base model. DmM, SWR, and GAM show the same tendency, and it can be said that this structure is stable. It is seen that in the base model, the relationship between $FS$ and unit resale prices can be estimated as a monotonically increasing function, and inaccurate prices can be obtained unless the nonlinearity is taken into account. The theoretical values for smoothing terms in GAM were normalized so that their total was 0.

![Diagram showing the relationship between floor space ($FS$) and unit resale prices](image)

**Figure 2.** Relationship between floor space ($FS$) and unit resale prices


Figure 3. Relationship between floor space ($FS$) and unit resale prices 2: GAM

This tendency of the relationship is considered to be attributable to differences in the characteristics and thickness of the market. First, condominiums of approximately $20 \text{ m}^2$ for single-person households are often purchased as investments, whereas purchasers are increasingly likely to be the persons living in condominiums as condominium size increases. Also, regarding construction costs, equipment such as kitchen and restroom equipment, the construction cost of which is significant, affects the area per unit price. Hence, construction cost per unit tends to increase as the area decreases. For these reasons, it is thought that the unit resale prices of condominiums with small areas tend to become higher.

The average $FS$ is $61.82 \text{ m}^2$. Condominiums of approximately $55$ to $70 \text{ m}^2$ for families are supplied in large numbers in Japan (Fig. 10), and the market becomes thinner as the condominium area increases. Because of this, as the area of a condominium increases, a premium is expected to be placed on units having $FS$ above a certain value.
5.2. Relationship between number of years since construction (Age) and unit prices

Next, attention was paid to the relationship between Age and unit resale prices. It was indicated that in the relationships between price reduction based on the values estimated using the base model formulated as a simple linear structure and using the three models, DmM, SWR, and GAM, greater discrepancies were observed in the price as the age departed from the average value (16.51 years). Regarding DmM, SWR, and GAM, we found that prices increased from around 12 years from construction and then decreased after 23 years. Reasons for the above trend may be that large-scale repair is first required 10 years after construction. In addition, similar large-scale repair is required around 20 years after construction, that is, 10 years after the first repair. Because the rate of depreciation is particularly rapid from 10 to 20 years after construction and because the building price value subsequently diminishes, the ratio of the land value increases, so the depreciation proportion is expected to decrease.

A comparison of the results obtained using the linear model with those obtained using DmM, SWR, and GAM showed that the price structure yielded by the linear model was greatly different from those given by DmM, SWR, and GAM and that, particularly in the case of the base model, as the price structure departed from the average price, greater discrepancy was observed.
Figure 4. Relationships between the number of years since construction (Age) and unit resale prices 1

Figure 5. Relationship between the number of years since construction (Age) and unit resale prices 2: GAM
5.3. Relationship between time to nearest station (TS) and unit resale prices

Analysis of the relationship between TS and unit resale prices showed that the price gradient increased slightly when TS was more than 12 minutes, and that it declined rapidly when TS was more than 17 minutes.

Figure 6. Relationship between time to nearest station (TS) and unit resale prices 1

Figure 7. Relationship between time to nearest station (TS) and unit resale prices 2: GAM
The above tendency was also observed in the cases of DmM, SWR, and GAM. Hence, when estimation was made using a linear model, price differences increased when secondhand condominiums were further from the nearest station. First, prospective buyers selecting a condominium site showed a high preference for transportation convenience, and when $TS$ was more than 10 minutes, the price declined. Because the analysis was made only for the walking time in this study, it is seen that there is a limiting point when $TS$ is more than 10 minutes. A greater price decline is observed when $TS$ exceeds 17 minutes. This is the limit for the perceived accessibility to the nearest station on foot, and if $TS$ exceeds this limit, access to the nearest station is thought to be by an alternative means such as bicycle, bus, or car.

5.4. Relationship between travel time to central business district ($TT$) and unit prices

Analysis of the relationship between $TT$ and unit resale prices showed that there was only one structural-change point, when $TT$ was more than 15 minutes, at which secondhand condominium prices declined rapidly. A similar tendency was also detected using DmM ($\beta = 1$) and SWR. In other words, secondhand condominium resale price levels did not differ from each other within the area of about 10 minutes from any one of the seven stations set as central business districts, but upon reaching a travel time of about 15 minutes, the prices declined. When using DmM, SWR, and GAM, but not the linear model, secondhand condominium resale price levels are estimated to increase slightly as the travel time increases up to about 10 minutes. It seems more appropriate to consider this as no price change, rather than an increased price, because of the great impact of estimation errors resulting from little change in price levels. However, secondhand condominium prices declined when $TT$ increased by more than 15 minutes.
The above analysis indicates that if an estimation were to be made using the linear model, secondhand condominium resale price levels would appear to decline more slowly away from the central business district, and for condominiums farther from the central business district, greater errors would arise. In GAM, which was affected by the large number of 10-minute samples, it was implicitly estimated that secondhand condominium prices first increased, then declined rapidly, then increased when $TT$ was 10 to 15 minutes, and ultimately declined again. This suggests that the use of GAM may result in errors if the data used for GAM distribution are discontinuous.

Figure 8. Relationship between travel time to central business district ($TT$) and unit resale prices
Figure 9. Relationship between travel time to central business district (TT) and unit resale prices 2: GAM

Table 7. Comparison estimated parameter between Base Model and GAM

<table>
<thead>
<tr>
<th>Method of Estimation</th>
<th>Method of Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>GAM</td>
</tr>
</tbody>
</table>

| Dependent Variable  |                      |
| RP: Resale Price of Condominiums (in log) |

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>3.931</td>
<td>155.275</td>
<td>3.475</td>
<td>299.292</td>
</tr>
<tr>
<td>FS: Floor space</td>
<td>0.047</td>
<td>8.984</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age: Age of building</td>
<td>-0.188</td>
<td>-96.379</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WT: Walk Time to the nearest station</td>
<td>-0.054</td>
<td>-21.510</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TT: Travel Time to CBD</td>
<td>-0.017</td>
<td>-5.537</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BS: Balcony space</td>
<td>0.032</td>
<td>4.471</td>
<td>0.008</td>
<td>3.198</td>
</tr>
<tr>
<td>NU: Number of units</td>
<td>0.020</td>
<td>10.190</td>
<td>0.032</td>
<td>17.477</td>
</tr>
<tr>
<td>RT: Market reservation time</td>
<td>-0.006</td>
<td>-3.331</td>
<td>-0.007</td>
<td>-4.536</td>
</tr>
<tr>
<td>Property Characteristics (in log)</td>
<td>Coefficient</td>
<td>t-value</td>
<td>Coefficient</td>
<td>t-value</td>
</tr>
<tr>
<td>Property Characteristics (dummy variables)</td>
<td>Coefficient</td>
<td>t-value</td>
<td>Coefficient</td>
<td>t-value</td>
</tr>
<tr>
<td>FF: First Floor Dummy</td>
<td>-0.034</td>
<td>-6.198</td>
<td>-0.043</td>
<td>-8.412</td>
</tr>
<tr>
<td>HF: Highest floor dummy</td>
<td>0.054</td>
<td>5.365</td>
<td>0.054</td>
<td>5.892</td>
</tr>
<tr>
<td>Ferroconcrete dummy</td>
<td>-0.012</td>
<td>-3.226</td>
<td>-0.016</td>
<td>-4.733</td>
</tr>
<tr>
<td>South-facing dummy</td>
<td>0.003</td>
<td>0.965</td>
<td>0.008</td>
<td>2.525</td>
</tr>
<tr>
<td>Ward (city) Dummy</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>RDi (i=0,...,I)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Railway/Subway Line Dummy</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>LDj (j=0,...,J)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDj (k=0,...,K)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DM2005Q2</td>
<td>0.002</td>
<td>0.351</td>
<td>-0.001</td>
<td>-0.358</td>
</tr>
<tr>
<td>DM2005Q3</td>
<td>0.014</td>
<td>3.271</td>
<td>0.014</td>
<td>3.584</td>
</tr>
<tr>
<td>DM2005Q4</td>
<td>0.022</td>
<td>5.037</td>
<td>0.022</td>
<td>5.394</td>
</tr>
<tr>
<td>Adjusted R square</td>
<td>0.775</td>
<td></td>
<td>0.810</td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td>9,682</td>
<td></td>
<td>9,682</td>
<td></td>
</tr>
</tbody>
</table>
6. Conclusions

In this study, we estimated the price structure of secondhand condominiums in the 23 wards of Tokyo by adopting a hedonic approach.

The secondhand condominium prices were explained in terms of condominium-oriented attributes, that is, building attributes including $FS$, $Age$ and the direction in which windows and the like face, and location attributes including $TS$, $TT$, $RD$ (which represents differences in the grade of areas along a railway line) and $LD$ (which represents grade differences between wards).

Attention was focused on the four main factors that affect secondhand condominium prices, $FS$, $Age$, $TS$, and $TT$. The relationship between each of these factors and secondhand condominium unit resale prices was assessed using a nonparametric model (DmM). The power of explanation was improved, and the determination coefficient adjusted for the degrees of freedom was 0.819 (0.775 for the base model). All four main factors, i.e., $FS$, $Age$, $TS$, and $TT$, were found to have nonlinear structures.

Estimation was also performed using SWR with AIC as an assessment index. It was found that there were nonlinear points at 40 and 90 m² in the relationship of prices with $FS$, at 12 and 23 years in the relationship with $Age$, at 12 and 17 minutes in the relationship with $TS$, and at 11 and 15 minutes in the relationship with $TT$.

Using the above analytical results, a structural-change test was conducted using the F-test. It was confirmed that structural changes occurred in relationships of price with $FS$, $Age$, and $TS$ at the two points that had been detected earlier (at a significance level of 10%) and that a structural change was observed only at 15 minutes but not at 11 minutes for $TT$.

On the basis of the above findings, on the assumption that there were two structural-change points in $FS$, $Age$, and $TS$ and that there was one structural change point in $TT$, function
estimations coupled with constant-term dummies and cross-terms were carried out. The determination coefficient adjusted for the degrees of freedom was improved to 0.817, similar to that for DmM.

When further estimation was performed using the GAM model, the power of explanation was improved, with a determination coefficient adjusted for the degrees of freedom becoming 0.819. The nonlinearity of the relationships of $FS$, $Age$, $TS$, and $TT$ with secondhand condominium resale prices was verified, as was the case for DmM and SWR. When the effectiveness of smoothing in the F-test was compared with the case of nonsmoothing, it was confirmed that a significant improvement of the model was obtained upon smoothing.

A comparison of the results obtained using the base model and the independent models, DmM, SWR, and GAM, showed that DmM, SWR, and GAM estimated a similar nonlinearity of the secondhand condominium resale price structure. In other words, it was indicated that if the secondhand condominium resale price structure were estimated using a linear model, there would be a region in which a large estimation error would arise. Because DmM, SWR, and GAM showed almost the same nonlinearity of the secondhand condominium resale price structure for $FS$, $Age$, and $TS$ and an improved power of explanation in terms of the determination coefficient adjusted for the degrees of freedom, SWR and GAM were demonstrated to be very practical and effective estimation methods. With regard to $TT$, however, DmM and SWR yielded similar estimates of the secondhand condominium resale price structure, but GAM yielded a different structure.

The results of formulating a secondhand condominium price function on the basis of the above-mentioned series of analytical results demonstrated/verified the ability of function estimation to produce results with a certain degree of accuracy. However, when a secondhand condominium resale price function was formulated for the Tokyo metropolitan district, it was difficult for the
simple linear model to yield accurate relationships between the main factors, \( FS, \ Age, \ TS, \) and \( TT \) and secondhand condominium unit resale prices, and these relationships were found to have nonlinear structures.

Figure10. Histogram of Main Variables
References


