Change in house price structure with time and housing price index

-Centered around the approach to the problem of structural change-

Chihiro SHIMIZU* Hideoki TAKATSUJI** Hiroya ONO** Kiyohiko G. NISHIMURA***

Abstract

The purpose of this study is to estimate the hedonic price index of secondhand condominiums while taking into account seasonal sample selection bias and structural changes, using the 23 wards of Tokyo as a subject. When housing price indices are estimated using a hedonic price model, the problem of temporal sample selection bias in addition to changes in the housing market structure should be considered. We propose an overlapping-period hedonic model (OPHM: this model was proposed by Ono, *et al* first), which can accommodate seasonal sample selection bias and structural changes. In addition, we estimate housing price indices for the 23 wards of Tokyo from 1986 through 2006, and demonstrate biases in price indices because of differences in the functions used in the models. Results of the estimation using the OPHM demonstrate that the structure of the housing market changes with time, and these changes occur continuously with time. It is also demonstrated that structurally restricted indices that do not account for structural changes involve a large time lag compared with indices that do account for structural changes during periods with significant price fluctuations. This study proposes a method of estimating hedonic housing price indices under the conditions of successively added data and structural changes.

Key Words: Hedonic housing price index, seasonal sample selection bias, (un)restricted hedonic model, overlapping-period hedonic model

JEL Code: C43 - Index Numbers and Aggregation, R31 - Housing Supply and Markets.

====================================

* Visiting Associate Professor, Center for Spatial Information Science (CSIS), University of Tokyo
Associate Professor, The International School of Economics and Business Administration, Reitaku University
2-1-1 Hikarigaoka, Kashiwa-Shi, Chiba, 277-8686 Japan
Tel. +81-(0)4-7173-3439, Fax. +81-(0)4-7173-1100
e-mail: cshimizu@reitaku-u.ac.jp
** Professor , The International School of Economics and Business Administration, Reitaku University
2-1-1 Hikarigaoka, Kashiwa-Shi, Chiba, 277-8686 Japan

***Member of the Policy Board, Bank of Japan
2-2-1 Hongoku-cho, Chuo-Ku, Tokyo 103-8660, Japan

1.Objectives of the study

The specifications and facilities of each house are different from each other in varying degrees, so there are no two houses of identical quality. Even when the specifications and facilities are identical, if the age of the building differs, the degree of deterioration differs accordingly, so that the houses are not identical. In other words, houses have "particularity with few equivalents". In addition to such a problem, the quality of houses (in particular, condominiums) changes with time owing to fairly rapid technological progress. Such characteristics are particularly evident in the housing market of Japan compared with the United States and other countries (Shimizu, Nishimura and Asami, 2004).

There are two approaches in constructing a housing price index that takes into account issues resulting from the above particularity with few equivalents and changes in quality: they are hedonic price model and the repeat-sales method. In the current study, we use hedonic price model to estimate the price indices of the secondhand condominium market in the 23 wards of Tokyo.

When the repeat-sales method and hedonic price model are compared, the following problems are noted for the respective methods.

In the repeat-sales method, the following two problems are noted: (i) a sample selection bias issue, that is, houses that are repeatedly on sale have different characteristics from the houses traded in the market (so-called lemons) (Clapp and Giaccotto, 1992), and (ii) accommodating qualitative changes and structural changes because of the assumption that there are no changes in the property characteristics and their parameters during the transaction period in the repeat-sales method (Case and Shiller, 1987, 1989; Clapp and Giaccotto, 1992, 1998, 1999; Goodman and Thibodeau, 1998; Case, Pollakowski and Wachter, 1991).

Furthermore, the estimation of housing price indices by hedonic price model has the following two major problems: (iii) the occurrence of bias in housing price indices because of the difficulty in collecting all the variables required for the estimation of functions as well as because of the presence of unobservable factors such as environmental variables (see Case and Quigley, 1991; Clapp, 2003), and (iv) a structural change issue, that is, it is necessary to accommodate changes in the house price structure, because the housing market is examined over a long period of time (Case, Pollakowski and Wachter, 1991; Clapp, Giaccotto and Tirtiroglu, 1991; Clapp and Giaccotto, 1992, 1998; Shimizu and Nishimura, 2006,2007). However, while the two approaches have problems in terms of estimation, it can also be shown that as the analysis period increases, the difference between the hedonic price index and the price index evaluated by the repeat-sales method decreases (Clapp and Giaccotto, 1998, 1999).

With regard to problem (i) of the repeat-sales method, because not all the transaction data are collected and estimated, the problem of sample selection bias also exists for hedonic price model , although its level is relatively low. Problems (ii) and (iv) occur in both the repeat-sales method and hedonic price model when housing price indices over a long period of time are estimated (i.e., there is a problem of coping with structural changes).

Regarding problem (iii), problems relating to the control of unobservable environmental variables can be avoided in the repeat-sales method (Case and Quigley, 1991; Case and Shiller, 1987, 1989; Thibodeau, 1997). Furthermore, the calculation procedure is simpler, and hence the calculation load is smaller in the repeat-sales method than in hedonic price model . Therefore, it has been considered that, at a glance, the repeat-sales method is more practical (Bourassa, Hoesli and Sun, 2006).

However, because the fluidity of housing markets is considerably lower (i.e., the market is thinner) in Japan than in the United States and other countries, and because institutional restrictions strongly suppress reselling within a short period, in accordance with the law based on the National Land Use Plan, problems with repeat-sales sample selection bias unique to Japan will still occur. If the repeat-sales method is to be applied in Japan, such a sample selection bias will be an extremely large problem, and in addition to that, the estimation of housing prices with high renewal frequency is impossible because of the small number of samples, and the estimation of indices in a limited area is difficult; consequently, the repeat-sales method is not a very practical method.

Under such circumstances, the importance of estimating hedonic housing price indices with high accuracy while solving the above problems involved in hedonic price model is extremely high in Japan. Therefore, in this study, we focus on the greatest problem involved in hedonic price model, which relates to changes in the market structure. Problem (iii) will be discussed in another report (according to Clapp (2003), unobservable variables are handled by adding coordinate data).

We start with the estimation of a structurally restricted hedonic model (hereafter, also referred to as the RHM) under the assumption of no changes in the market structure and a structurally unrestricted hedonic model (hereafter, also referred to as the URHM) under the assumption that the structure changes in each period (Case, Pollakowski and Wachter, 1991; Clapp, Giaccotto and Tirtiroglu, 1991).

In Section 2, the structures of the structurally restricted hedonic housing price index (hereafter, also referred to as the RHI) estimated using the restricted hedonic model and of the structurally unrestricted hedonic housing price index (hereafter, also referred to as the URHI) estimated using the unrestricted hedonic model are explained together with the repeat-sales index. This is aimed at clarifying the characteristics of hedonic price model in terms of estimation, by means of comparison with the repeat-sales method. Then, we propose a new housing price index taking into account structural changes and seasonal sample selection bias; the overlapping-period hedonic housing index (hereafter, also referred to as the OPHI). Data are explained in Section 3, and RHI, URHI and OPHI are estimated for the secondhand condominium market in the 23 wards of Tokyo, and the estimated housing price indices are evaluated in Section 4.

The results show temporal changes in the house price structure in the secondhand condominium market, and in particular, regarding the floor space, the sign is reversed during some periods. Therefore, it is necessary to estimate the price index taking into account structural changes. In addition, because housing transactions in Japan tend to be concentrated in periods when large numbers of people move, the number of samples varies considerably with the season; accordingly, seasonal sample selection bias should be considered. To accommodate changes in the market structure and the problem of seasonal sample selection bias, it has been demonstrated that OPHM with an overlapped estimate period τ of 12 months is effective.

2.Changes in market structure and housing price indices

2.1.Structurally restricted hedonic housing price index (RHI) and structurally unrestricted hedonic housing price index (URHI)

Regarding the estimation of quality-adjusted housing price indices, the hedonic estimation method and the repeat-sales estimation method can be used. The price indices of the hedonic estimation method include the structurally restricted price index and the structurally unrestricted price index. We summarize these estimation methods, and clarify the characteristics of the hedonic price indices in terms of estimation in comparison with the repeat-sales method.

2.1.1.Structurally restricted hedonic housing price index: RHI

Assume that we have data for the house price and residential property characteristics, which are pooled for all the periods $t = 1, 2, \ldots, T$, and that the number of data samples in each period is n_t. A house price estimation model that can be used to obtain a structurally restricted price index is given as follows.

$$\ln P_{it} = \sum_{k=1}^{K} \beta_k X_{ikt} + \sum_{s=1}^{\tau} \delta_s D_s + \varepsilon_{it} \tag{1}$$

$t = 1, 2, \ldots, T$.

$i = 1, 2, \ldots, n_t$ (designates ith data among the n_t data samples in period t).

P_{it} = price of house i in period t (designates ith data among the data in period t, instead of designating the same house i over each of the t periods).

β_k = parameter of residential property characteristic k.

X_{ikt} = value of property characteristic k of house i in period t.

δ_s = parameter of the time dummy variable in period s.

D_s: when $s = 1$, this takes a constant value of 1 (constant term). When $2 \le s \le T$, this is a time dummy variable, and it takes a value of 1 when s = t and a value of 0 otherwise.

ε_{it} = random disturbance term.

This model is called the structurally restricted hedonic model (RHM) because it assumes that the regression coefficient β_k of the house-price determining factor X_{ikt} is constant throughout all the periods. From this, the RHI is obtained as follows. The estimated price $\hat{P}_t$ of a house with residential property characteristic values $\{X_k\}$ ($k = 1, 2, \ldots, K$) in period t ($t = 1, 2, \ldots, T$) is given as follows.

$$\ln \hat{P}_t = \sum_{k=1}^{K} \hat{\beta}_k X_k + \hat{\delta}_1 + \hat{\delta}_t \tag{2}$$

$$\ln \hat{P}_1 = \sum_{k=1}^{K} \hat{\beta}_k X_k + \hat{\delta}_1 \tag{3}$$

Here, $\hat{\beta}_k, \hat{\delta}_1, \hat{\delta}_t$ are estimated values of the parameters. Accordingly, the housing price index $\hat{P}_t / \hat{P}_1$ in period t, where the house price in period $t = 1$ is used as the reference, is obtained as follows.

$$\ln(\hat{P}_t / \hat{P}_1) = \hat{\delta}_t \tag{4}$$

In addition, the change in the price index from period $t-1$ to period t can be expressed as follows.

$$\ln(\hat{P}_t / \hat{P}_{t-1}) = \hat{\delta}_t - \hat{\delta}_{t-1} \tag{5}$$

In this case, the price index is obtained under the assumption of specific residential property characteristic values $\{X_k\}$; however, as we can see in the above process, price indices are expressed using only time dummy variables without involving residential property characteristic values in the RHI.

2.1.2.Structurally unrestricted hedonic housing price index: URHI

Using similar data to those described above, a house price estimation model that can be used to obtain a URHI is given as follows.

$$\ln P_{it} = \sum_{k=1}^{K} \beta_{kt} X_{ikt} + \delta_t + \varepsilon_{it} \tag{6}$$

Here, no time dummy variables are used, and instead, the parameter β_{kt} and a constant term δ_t related to residential property characteristics are assumed to change in each period. Namely, because the model does not assume the restriction of constant parameters, it is called the URHM. From equation (6), the URHI is obtained as follows. The estimated price $\hat{P}_t$ of a house with residential property characteristic values $\{X_k\}$ ($k = 1, 2, \ldots, K$) in period t ($t = 1, 2, \ldots, T$) is given as follows.

$$\ln \hat{P}_t = \sum_{k=1}^{K} \hat{\beta}_{kt} X_k + \hat{\delta}_t \tag{7}$$

$$\ln \hat{P}_1 = \sum_{k=1}^{K} \hat{\beta}_{k1} X_k + \hat{\delta}_1 \tag{8}$$

Therefore, the housing price index $\hat{P}_t / \hat{P}_1$ in period t, where the house price in period $t = 1$ is used as the reference, is obtained as follows.

$$\ln(\hat{P}_t / \hat{P}_1) = \sum_{k=1}^{K} (\hat{\beta}_{kt} - \hat{\beta}_{k1}) X_k + (\hat{\delta}_t - \hat{\delta}_1) \tag{9}$$

In addition, the change in the price index from period $t-1$ to period t can be expressed as follows.

$$\ln(\hat{P}_t / \hat{P}_{t-1}) = \sum_{k=1}^{K} (\hat{\beta}_{kt} - \hat{\beta}_{k,t-1}) X_k + (\hat{\delta}_t - \hat{\delta}_{t-1}) \tag{10}$$

Thus, for the URHI, price indices are obtained for specific residential property characteristics. When specific residential property characteristics change, the price index changes accordingly.

2.1.3. Repeat-sales housing price index

Next, the repeat-sales method is summarized. For house h, it is assumed that its price is determined by the residential property characteristics and the time point of the transaction. It is also assumed that the residential property characteristics do not change with time, and the strength of their effect on price formation also does not change. Thus, the house price model in this case can be expressed as follows.

$$\ln P_{ht} = \sum_{k=1}^{K} \beta_k X_{hk} + \sum_{s=1}^{T} \delta_s D_s + \varepsilon_{ht} \tag{11}$$

P_{ht} is the price of house h in period t. Here, we assume that house h appears repeatedly in different periods. X_{hk} is the value of property characteristic k of house h, which does not change with time. Accordingly, we assume that parameter β_k of X_{hk} also does not change with time. D_s is a time dummy variable and equals 1 when $s = t$ (the period of the transaction) and 0 otherwise. Here, we assume that $D_1 = 1$ (a constant term). δ_s is the parameter of the time dummy variable. It is assumed that house h is subject to a transaction twice, in periods t_1 and t_2, during the estimation period of $t = 1, 2, \ldots, T$. The house prices in the periods of transaction can be expressed, using the above model, as follows.

$$\ln P_{ht_1} = \sum_{k=1}^{K} \beta_k X_{hk} + \delta_1 + \delta_{t_1} + \varepsilon_{ht_1} \tag{12}$$

$$\ln P_{ht_2} = \sum_{k=1}^{K} \beta_k X_{hk} + \delta_1 + \delta_{t_2} + \varepsilon_{ht_2} \tag{13}$$

From equations (12) and (13), the price change P_{ht_2} / P_{ht_1} is given as follows.

$$\ln(P_{ht_2} / P_{ht_1}) = \delta_{t_2} - \delta_{t_1} + \left(\varepsilon_{ht_2} - \varepsilon_{ht_1}\right) \tag{14}$$

Therefore, in this model, the price change is determined from the difference in the two time points of the transactions irrespective of residential property characteristics. We now formulate the model to estimate changes in the house price with respect to data collected for various transactions of houses at various time points. We obtain:

$$\ln(P_{ht_2} / P_{ht_1}) = \sum_{s=1}^{T} \delta_s D_s + \mu_h , \tag{15}$$

where μ_h is a random disturbance term and D_s is a time dummy variable, which takes a value of 1 at the second transaction ($t = t_2$), −1 at the first transaction ($t = t_1$), and 0 in other periods. In addition, parameter δ_s of each time dummy variable estimated using this model represents the price index of each period. This represents the typical repeat-sales model.

On the basis of the above summary of the indices and under the assumption that the market structure changes, the structurally restricted hedonic price index is problematic in that it has the restriction that parameters for X_{ki} are identical throughout all the periods. In the repeat-sales method, the same assumption that parameters for X_{ki} remain constant is adopted, and in addition, the very strong assumption that there are no changes in the property characteristics during the transaction period is adopted as well. The latter assumption is too strong for Japan. In practice, the values of houses change because of the extension and reconstruction, or renovation of buildings, and because of the development of physical damage accompanying the increasing age of buildings. In Japan, large-scale renovation is performed every couple of years, and the life of houses is short compared with that in Europe and the United States; therefore, house prices significantly decrease as the age of buildings increases. Moreover, because of the weak restrictions on city planning in Japan, it is extremely unrealistic to assume that there are no changes in the city environment or in building property characteristics.

In the URHM , although the restriction on parameters for X_{ki} can be eliminated, an assumption that the parameters including error terms are independent in each period is necessary. However, in a real market, it cannot be expected that the structure always changes randomly. In addition, a new issue arises. When housing price indices are set with a period of less than one year, i.e., quarterly or monthly, a problem of seasonal sample selection bias occurs. For example, in Japan there are periods with a large number of transactions, from January to March when many people move, and periods with a small number of transactions, such as July and August, and these seasonal changes in the number of transactions may affect and bias the price index. Thus, although the URHM can possibly accommodate structural changes, it disconnects the estimation from the continuity of the market conditions and generates the problem of seasonal sample selection bias.

2.2.Overlapping-period hedonic housing price index: OPHI

The URHI assumes that the market structure changes successively. Such a structural change of the market occurs as a result of various external shocks; it is considered that there is, in reality, a certain adjustment period before such a change penetrates into the market. Accordingly, regression coefficients should be regarded as changing successively rather than instantaneously. However, generally, the estimation of a model with structural changes is performed by dividing observation data into several periods with break points, then using the divided data of each period (for example, Ono *et al*,2004; Shimizu and Nishimura, 2006; Shimizu and Nishimura, 2007). Namely, the continuity of the observation data is disconnected at the break points. Therefore, it is rather difficult to use such an estimation method, under the assumption of the occurrence of successive structural changes, to determine regression coefficients allowing for successive changes. Instead, it may be more natural and desirable to estimate regression coefficients on the basis of a process of successive change by taking a certain period length τ as the estimation period, and by shifting this period, similar to the process of obtaining moving averages. This process can be formulated as follows.

Assuming that we have pooled data over the periods 1, 2, …T. With respect to some of these periods, i.e., a period length τ, we assume the following basic model.

$$\ln P_{it} = \sum_{k=1}^{K} \beta_k X_{ikt} + \sum_{s=1}^{\tau} \delta_s D_s + \varepsilon_{it} , \qquad (16)$$

where $t = 1, 2, \ldots, \tau$ (taking part of the entire pooled data consisting of 1, 2, …T periods, namely, taking a certain period length τ, the periods within this range are numbered from 1 to τ).

$i = 1, 2, \ldots, n_t$ (ith datum among n_t items of data in period t).

P_{it}: price of house i in period t.

β_k: parameter of residential property characteristic k; β_k is assumed not to change within the period length τ.

X_{kit}: value of property characteristic k of house i in period t.

δ_s: parameter of the time dummy variable in period s.

D_s: when $s = 1$, this takes a constant value of 1 (constant term). When $2 \le s \le \tau$, this is a time dummy variable and takes a value of 1 when $s = t$ and a value of 0 otherwise.

ε_{it}: random disturbance term.

In addition, we express a period with length τ starting from period r among periods 1, 2, …T, as $[r, r + \tau - 1]$. Then, our estimation method is obtained by applying the above basic model to periods $[1, \tau]$, $[2, \tau + 1]$, …, $[r, r + \tau - 1]$, …,$[T - \tau + 1, T]$ successively. From this, successive changes of the market structure can be reflected in changes in the parameters. We call this model the OPHM, and the period length τ, the overlapped estimate period length.

OPHM is a RHM with respect to a certain period length τ. Accordingly, the parameter of the time dummy variable represents the price index of each period with the starting period of length τ as the reference. Thus, price indices can be obtained directly from the basic model within the period length τ. With the OPHM, models for all the periods are estimated by successively shifting the period length τ by one period. Here, the problem remains of how to connect the price indices obtained by the estimation in each period length τ to construct the price index for all periods. Our method is as follows.

We designate the housing price index throughout all the periods as q_r. This represents the price index of period r among periods 1, 2, …T. We designate the reference period as period 1, and assume that $q_1 = 0$. We also designate a parameter of the time dummy variable obtained by applying the basic model to the data for the period length τ starting from period r among periods 1, 2, …T, i.e., $[r, r + \tau - 1]$, as

$\hat{\delta}_1^{(r)}, \hat{\delta}_2^{(r)}, \cdots, \hat{\delta}_\tau^{(r)}$ by explicitly expressing the period r.

The procedure to obtain the housing price index q_r is as follows.

(Set 1)

The basic model is applied to the first $[1, \tau]$ period to obtain the parameter of the time dummy variable.

$$\hat{\delta}_1^{(1)}, \hat{\delta}_2^{(1)}, \cdots, \hat{\delta}_\tau^{(1)}$$

Using these parameters, we define the price index q_r ($r = 1, 2, \ldots, \tau$) for the $[1, \tau]$ period as follows.

$$\begin{aligned} q_1 &= 0 \\ q_2 &= \hat{\delta}_2^{(1)} \\ q_3 &= \hat{\delta}_3^{(1)} \\ &\cdots \\ q_\tau &= \hat{\delta}_\tau^{(1)} \end{aligned} \tag{17}$$

(Set 2)

To obtain the next price index $q_{\tau+1}$ under the assumption that those up to q_τ have been determined as described above, an estimated amount, which is considered to be the change from q_τ to $q_{\tau+1}$, is added to q_τ. We consider that this estimated amount is based on the following parameters:

$$\hat{\delta}_1^{(2)}, \hat{\delta}_2^{(2)}, \cdots, \hat{\delta}_\tau^{(2)}, \tag{18}$$

which are determined by the estimation using the basic model for the next $[2, \tau + 1]$ period, as follows.

$$\delta_{\tau}^{(2)} - \delta_{\tau-1}^{(2)} \tag{19}$$

Accordingly, $q_{\tau+1}$ is defined as follows.

$$q_{\tau+1} = q_{\tau} + (\delta_{\tau}^{(2)} - \delta_{\tau-1}^{(2)}) \tag{20}$$

(Set 3)

Similarly, to obtain the next price index $q_{\tau+r-1}$ under the assumption that those up to $q_{\tau+r-2}$ have been determined, an estimated amount, which is considered to be the change from $q_{\tau+r-2}$ to $q_{\tau+r-1}$ is added to $q_{\tau+r-2}$. Accordingly, on the basis of the parameters determined by the estimation of the basic model for the $[r,\ \tau + r - 1]$ period,

$$\hat{\delta}_1^{(r)}, \hat{\delta}_2^{(r)}, \cdots, \hat{\delta}_{\tau}^{(r)}, \tag{21}$$

the estimated price index is defined as follows.

$$q_{\tau+\tau-1} = q_{\tau+\tau-2} + (\delta_{\tau}^{(r)} - \delta_{\tau-1}^{(r)}) \tag{22}$$

Thus, we can obtain the price indices by OPHM for all periods.

Here, we note one point regarding the estimation of the basic model (equation (16)). With respect to the random disturbance term ε_{it}, when we assume:

$$Var(\varepsilon_{it}) = \sigma_t^2, \tag{23}$$

it has been confirmed by our previous analysis that:

$$\sigma_i^2 \neq \sigma_j^2 \quad (i \neq j) \tag{24}$$

holds. Namely, heterogeneity of the variance is observed. Therefore, the basic model (equation (16)) is reestimated using feasible generalized least squares (FGLS). That is, $\hat{\sigma}_t^2$ is obtained from the residual upon estimation using the basic model (equation (16)), and then the parameter is estimated using the following.

$$(\ln P_{it})/\hat{\sigma}_t = \sum_{k=1}^{K} \beta_k (X_{ikt}/\hat{\sigma}_t) + \sum_{s=1}^{\tau} \delta_s (D_s/\hat{\sigma}_t) + (\varepsilon_{it}/\hat{\sigma}_t)$$

(25)

2.3.Setup of estimation model

In this study, the secondhand condominium market in the 23 wards of Tokyo is used as the analytical subject. The basic equation of the RHM is as follows.

$$\log RP/FS = a_0 + \sum_h a_{1h} \log X_h + \sum_i a_{2i} \log Z_i + \sum_j a_{3j} \cdot LD_j + \sum_k a_{4k} \cdot RD_k + \sum_l a_{5l} \cdot TD_l + \varepsilon$$

(26)

RP: Resale price of condominium (yen)

X_h: Main variables

- *FS*: Floor space (square meters)
- *Age*: Age of building (months)
- *TS*: Time to the nearest station (minutes)
- *TT*: Travel time to central business district

Z_i: Other variables

- *BS*: Balcony space (square meters)
- *NU*: Number of units
- *BC*: Other building characteristics

RT: Market reservation time (weeks)

LD_j: Location (ward) dummy (j = 0 … J)

RD_k: Railway line dummy (k = 0 … J)

TD_l: Time dummy (l = 0 … K)

Residential property characteristics as explanatory variables include floor space (*FS*), time to nearest station (*TS*), travel time (*TT*) to the central business district (CBD), age of building (*Age*), area of balcony (*BS*) and other building property characteristics (Z_h), as well as a railway dummy variable (RD_j) and location (ward) dummy variable (LD_k) as location factors. In addition, we have a time dummy variable (TD_k). The regression coefficient of this time dummy variable (a_{12k}) represents the secondhand condominium price index. In the RHM, the estimation is performed using this equation with pooled data in the periods $t = 1 \cdots T$.

In contrast, the basic equation of the URHM is given by equation (27) below; this is the equation of the RHM from which the time dummy variable is excluded. Estimation is performed for each period using the data of that period (period *t*). Setting the period length to one month, a model is estimated for all periods. Then, using the estimated models and the assumption of houses having the same quality, the price of the houses in each period is estimated and its change-over time is examined.

$$\log RP_t / FS_t = a_0 + \sum_h a_{1th} \log X_{th} + \sum_i a_{2ti} \log Z_{ti} + \sum_j a_{3j} \cdot LD_j + \sum_k a_{2tk} \cdot RD_{tk} + \varepsilon$$

$$t = 1 \cdots\cdots\cdots T \tag{27}$$

Using the OPHM, the RHM of equation (26) is adopted for the estimation period length τ.

3.Data

3.1.Secondhand condominium price data

The subject of the analysis is the 23 wards of Tokyo metropolitan area (621 square kilometers), and the analysis period is approximately 20 years between January 1986 and September 2006.

As the main information source, we used the prices of secondhand condominiums published in Residential Information Weekly (or Shukan Jyutaku Joho in Japanese) published by RECRUIT, Co. This magazine provides information on the quality and asking price of listed properties on a weekly basis, and includes the historical price data of individual properties from the time they are placed in the magazine for the first time until they are removed because of sale or other reasons. There are three items of information regarding the price: i) the initial asking price (first offer price) upon appearance in the market, ii) the price upon removal from the magazine (estimated purchase price: first bid price), and iii) the transaction price, which are collected as a sample. The first asking price represents the seller's desired price rather than the market value. In contrast, some of the transaction prices may partly reflect the specific situations of individual transactions such as the desire for a quick sale or hasty purchase.

Therefore, among the information published in Residential Information Weekly, we decided to use the price when the listing of the house is removed from the magazine upon the conclusion of the contract as the explanatory variable in the model. The price at the time of removal from the magazine is the first bid price offered by a prospective buyer, such a bid is offered through the process in which several particulars of quality and price are disclosed to the market via the magazine, and the price is decreased until the buyer responds to that information. Thus, the price we used indicates the upper range of possible bid prices and can be regarded as a competitive market price that is relatively free from individual specific conditions associated with transactions.

3.2.Data regarding house quality

In the condominium market, while condominiums with light-gauge steel structures are also included, transactions are mainly for condominiums with reinforced concrete (RC) structures or steel-reinforced concrete (SRC) structures; therefore, we used condominiums with RC or SRC structures as the subject of our study. A list of data analyzed is shown in table 1.

The transport accessibility of each condominium location is represented by the *TT* to the CBD and the

TS. The former is measured in the following way. First, we defined the CBD. The Tokyo Metropolitan area is composed of the 23 wards of Tokyo as its center, with a dense railway network developed therein. We designated seven terminal stations, which include six on the Yamanote Line, Tokyo, Shinagawa, Shibuya, Shinjuku, Ikebukuro, and Ueno, as well as Otemachi as the central station of the Tokyo Metro (Teito Rapid Transit Authority). Then, we investigated average travel times during the day from each station to the seven terminal stations, and set the minimum value as the *TT* to the CBD for that station. By considering travel times to multiple terminal stations, the timesaving effect of a new railway development on the whole transport network can be embedded into the model. When a new line is developed or a new station is constructed, or when a timetable is changed, this index will change. Therefore, travel times are renewed once every half year (April and October).

Regarding *TS*, different means of transportation are available. There are three transportation means: on foot, by bus and by car. However, because a very dense transportation network is established in the 23 wards of Tokyo, and many condominiums are built in transportation-convenient areas, condominiums only within walking distance or bus-transportation distance are included in the analysis data. Therefore, any difference in the transportation distance between the former and latter is controlled for using the bus dummy variable (*BD*). In addition, the walking time (in minutes) is recorded when the condominium is within walking distance, and the walking time from the condominium to the bus stop and the on-board time from the bus stop to the nearest station (in minutes) are recorded in the case of condominiums in a bus-transportation area. The *TS* is defined as; (walking time to nearest station) + (walking time to the bus stop) + (on-board time from the bus stop to the nearest station). Then, with respect to the bus-transportation area, the cross term of the constant dummy variable with the *TS* is incorporated in the BD. Details of the model are described below.

The sale price of each property is also affected by the fluidity and thickness of the market. The time spent until the contract is concluded is considered to be affected by the period and location and by the level of activity of transactions in the market. We explain such market factors using variables such as market reservation time (*RT*). This is the time period between the date when a house is placed on the market by a seller, and the time period when a buyer appears. Properties with a long *RT* are regarded as those having a price higher than the equilibrium price, or those in a thin market. Conversely, properties with a short *RT* are regarded as those in a market with high fluidity or those having a price close to or lower than the equilibrium price. For our purposes, the *RT* is defined as the time period from when the property is listed in the magazine for the first time until it is removed from the magazine.

We also identified some quantitative measures representing building property characteristics. They are *FS*, *Age*, *BS* and the number of units (*NU*). *Age* is the time period from its construction to the time period when the contract is concluded. *NU* is regarded as a proxy variable for the grade of the entire condominium and the quality of the common space. In addition, we created other dummy variables: the ground-floor dummy variable is used because the price of a ground-floor property is expected to be lower than that of higher-floor properties, and the highest-floor dummy variable is used because the price is expected to again be higher on the top floor. In terms of the direction of properties, such as whether their opening parts are south facing or not, we define a south-facing dummy variable (*SD*). Furthermore, we define a ferroconcrete structure dummy variable (*FD*) to incorporate the difference in structural strength.

The above variables are all related to the location or building characteristics of condominiums. It is reasonable to assume that other regional factors may also affect the house price. Therefore, we set an administrative-area dummy variable, *LD*, to reflect differences in the quality of public services, as well as the "*Ji-Gurai*" or area rank. Furthermore, the railway dummy variable, *RD*, represents the price structure for condominiums along railway lines, because most Japanese residential developments have been carried out along railway lines.

Finally, the time dummy variable (*TD*) is used to control for differences in the prices between different time periods.The observation data consist of 211,179 samples collected between January 1986 and September 2006.

Table 1. List of analyzed data.

Symbols	Variables	Contents	Unit
FS	Floor space/ square meters	Floor space.	m^2
AGE	Age of Building :Number of years since construction	Period between the date when the data is deleted from the magazine and the date of construction of the building.	month
TS	Time to nearest station	Time distance to the nearest station (Time by Walk or Bus).	minute
TT	Travel Time to central business district	Minimum of railway riding time in daytime to Terminal 7 stations in 2005*.	minute
BS	Balcony space/ square meters	Balcony space.	m^2
NU	Number of units	Total units of the condominium.	unit
RT	Market reservation time	Period between the date when the data appear in the magazine for the first time and the date of being deleted.	week
FD	First floor dummy	The property is on the ground floor 1, on other floors 0.	(0,1)
HF	Highest floor dummy	The property is on the top floor 1, on the other floors 0.	(0,1)
SD	South-facing dummy	Fenestrae facing south 1, other directions 0.	(0,1)
FD	Ferroconcrete dummy	Steel reinforced concrete frame structure 1, other structure 0.	(0,1)
LDj (j=0,…,J)	Location (Ward) dummy	*j th administrative district 1,* other district 0.	(0,1)
RDk (k=0,…,K)	Railway line dummy	*k th railway line 1,* other railway line 0.	(0,1)
TDl (l=0,…,L)	Time dummy (monthly)	*l th month 1,* other month 0.	(0,1)

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

3.3.Statistical distribution of secondhand condominium price data

Table 2 shows descriptive statistics of the major variables. The average resale price of a condominium is 39.04 million yen, the minimum value is 8.50 million yen, and the maximum value is 195.00 million yen, with a fairly large standard deviation of 23.48 million yen. The data include a wide range of condominiums from studio-apartment-class small properties to the so-called 100-million-yen-class large properties. The average unit price is approximately 0.7 million yen/m^2 with a right-skewed distribution.

Regarding the *FS*, the minimum value is 16.00 m^2, the maximum value is 134.99 m^2, and the average is 56.57 m^2, including all condominiums from single-person households to large-family condominiums.

Regarding the *Age*, the average value is 165 months (13.75 years), with a maximum value of 413 months (34.42 years). Because the history of condominiums in Japan is short, it is expected that this index will increase over time.

Regarding the *TS*, we only observed the distribution of data on the time axis; there are properties with a minimum value of 0 minutes that are located in front of a station. The maximum value is 32 minutes and the average value is 7.60 minutes. On average, while many properties are conveniently located, some are beyond walking distance. This indicates that, in general, convenience is emphasized in the construction of condominiums because of the required characteristics of condominiums.

Regarding the *TT* to the CBD, the average is 15 minutes with a maximum value of 30 minutes, indicating that most condominiums are concentrated in areas of high convenience.

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

Variables	Average	Standard Deviation	Minimum	Maximum
RP: Resale Price of Condominimum (10,000 Yen)	3,904.66	2,348.54	850.00	19,500.00
FS: Floor space (m²)	56.57	19.40	16.00	134.99
RP/FS	70.93	36.78	24.00	270.90
Age :Age of Building(months)	165.74	91.98	5.00	413.00
TS: Time to the nearest station: (minutes)	7.60	4.27	0.00	32.00
TT: Travel Time to Central Business District (minutes)	15.32	5.30	0.00	30.00
NU: The Number of Units	100.03	131.05	10.00	1149.00
RT: Market reservation time (week)	11.58	10.62	1.00	64.00

1986/01-2006/09 n=211,179

4.Estimation results

4.1.Estimation of RHI

The estimated RHI for the 23 wards of Tokyo is as follows.

$$\log RP/FS = \underset{(498.23)}{4.631} + \underset{(+10.81)}{0.0126} \cdot \log FS - \underset{(-337.38)}{0.189} \cdot \log Age - \underset{(-99.69)}{0.078} \cdot \log TS - \underset{(-36.21)}{0.117} \cdot \log TT + \underset{(40.90)}{0.019} \cdot \log NU$$

$$- \underset{(-13.140)}{0.276} \cdot BD + \underset{(6.970)}{0.058} \cdot (BD \times \log WT) - \underset{(-19.210)}{0.026} \cdot FF + \underset{(8.000)}{0.018} \cdot HF - \underset{(-10.150)}{0.097} FD + \underset{(10.790)}{0.0093} \cdot SD$$

$$+ \beta_{1h} \sum_h LD_h + \beta_{2i} \sum_i RD_i + \beta_{3j} \sum_j TD_j$$

Adjusted R-square: 0.837

Number of observations: 211,178

Because the coefficient of determination adjusted for the degrees of freedom is 0.837, the estimated model has a fairly high explanatory power (refer to Table 3 for details).

Because the data was pooled for sales between 1986 and 2006, we corrected the time point by forcibly introducing the *TD*, so that the structure of the secondhand condominium price was estimated using property characteristics specific to condominiums and the *RD*. Among the property characteristics specific to condominiums, *FS*, *BS* and *NU* have positive values, and *Age*, *TS*, and *TT* to the CBD are estimated with negative values.

First, regarding *FS*, the unit price was shown to increase with increasing floor space. A similar tendency was observed for *BS* and *NU*. This indicates that consumers show a strong preference for the floor space of each property as well as the floor space of the entire condominium.

As *Age* increases, we expect not only functional deterioration but also economic deterioration because of the improvement of facilities in newer condominiums. The results obtained showed that as *TS* and *TT* to the CBD increase, the convenience decreases because of the greater distance from populated areas, resulting in a decrease in the price.

Furthermore, the level of public service differs for each administrative ward, and there are broad differences in the residential environment depending on administrative cities and wards or railway line areas, which cannot be taken into consideration in our estimated function; therefore, these differences were estimated using the dummy variables.

Table 3. Estimation results of the RHM: 23 wards of Tokyo.

Method of Estimation
OLS
Dependent Variable
RP : Resale Price of Condominiums (in log)
Independent Variables

Property Characteristics (in log)	Coefficient	t-value
Constant	4.631	498.230
FS : Floor space	0.013	10.810
Age : Age of building	-0.190	-337.380
TS : Time to the nearest station	-0.078	-99.690
TT : Travel Time to CBD	-0.040	-36.210
NU : Number of units	0.019	40.900
RT : Market reservation time	0.014	32.530
Property Characteristics (dummy variables)	**Coefficient**	**t-value**
BD : Bus Dummy	-0.276	-13.140
TS × *BD*	0.059	6.970
FF : First Floor Dummy	-0.026	-19.210
HF :Highest floor dummy	0.018	8.000
FD :*Ferroconcrete dummy*	-0.010	-10.150
SD :*South-facing dummy*	0.009	10.790
Location (Ward) Dummy *LDj (j=0,…,J)*	**Coefficient**	**t-value**
Chiyoda	0.625	110.740
Chuo	0.347	82.770
minato	0.552	154.730
Shinjuku	0.407	115.620
Bunkyo	0.356	95.060
Taito	0.047	10.080
Koto	-0.030	-8.970
Shinagawa	0.315	86.020
Meguro	0.443	109.280
Ota	0.233	62.930
Setagaya	0.407	115.890
Shibuya	0.583	155.950
Nakano	0.284	65.620
Suginami	0.248	60.990
Toshima	0.243	61.220
Kita	0.092	17.620
Arakawa	-0.064	-14.580
Itabashi	0.007	1.960
Nerima	0.146	37.100
Adachi	-0.171	-43.080
Katsushika	-0.144	-38.390
Edogawa	-0.080	-21.890

Railway/Subway Line Dummy *LDk (k=0,…,K)*	Coefficient	t-value
Yamanote	0.033	4.236
Ginza	0.158	11.460
Marunouchi	0.056	5.556
Hibiya	0.085	9.039
Tozai	0.040	4.727
Chiyoda	0.067	7.858
Yurakucho	0.053	3.609
Hanzomon	-0.029	-2.621
ToeiAsakusa	-0.265	-2.420
ToeiShinjuku	-0.338	-10.244
Keihinkyukou	-0.214	-15.225
Keihintohoku	-0.265	-6.692
Ikegami	-0.089	-6.782
Tokyutoyoko	0.036	1.712
Setagaya	-0.091	-5.465
Odakyu	-0.025	-1.649
Inogashira	0.076	5.800
keio	0.032	2.361
Chuo	-0.045	-1.621
Seibushinjuku	-0.053	-4.024
Seibuikebukuro	0.040	2.455
Toubutojou	-0.126	-10.416
Saikyo	0.065	5.680
Takasaki	-0.063	-5.655
Toubuisezaki	-0.073	-2.694
Jouban	-0.111	-7.410
Soubu	-0.122	-5.927
Time Dummy *TDl (l=0,…,L)*	**Coefficient**	**t-value**
yes(see Figure)		

Adjusted R square= 0.837
Number of Observations= 211,179

4.2.Estimation of URHI

Next, we estimated the URHM. In accordance with the definition of equation (27), we divided the data into t periods (here, monthly) and estimated the structure of house prices. Regarding the price index, we estimated the prices of secondhand condominiums in each period by substituting the specific residential property characteristics common to all periods into the explanatory variables, and obtained the structurally unrestricted hedonic housing price indices relative to the reference period based on the estimated prices.

Table 4 shows the estimated regression coefficients of the major variables, and Fig. 1 shows changes in the number of samples and the coefficient of determination adjusted for the degrees of freedom.

The adjusted coefficient of determination decreased from 1986 through 1995, then increased from 1996. However, on the whole, it maintained an average of around 0.75, showing fairly good results.

The number of samples was approximately 500 per month from 1986 through 1989, which then increased significantly to an average value of 844. However, there is more than a threefold difference depending on the month. In each year, transactions are concentrated from January to March, which is the end of the fiscal year, when there are large movements of people in Japan, and the number of transactions significantly decreases around July and August, thus showing seasonal changes. However, there is no apparent correlation between the number of samples and adjusted coefficient of determination.

Next, we focused on the regression coefficients of the estimated model. Table 5 shows descriptive statistical values for the regression coefficients over 250 periods. Figs 2–6 show changes in the regression coefficients with time. All of the regression coefficients show major fluctuations in each period or every several periods. However, the fluctuations show a certain tendency, although not a gradual change with time. In addition, we can see that all the variables are around (higher or lower than) the regression coefficients estimated using the RHM.

Using the data in Table 5, we calculated the coefficients of variance (standard deviation/average): *FS* = 2.428, *Age* = –0.179, *TS* = –0.232, *TT* to CBD = –0.779. In other words, *FS* shows the largest change, including a change of the sign (+/–), followed by *TT* to the CBD, *TS*, and *Age*.

Table 4. Estimation results of the URHM : the 23 wards of Tokyo: 1986/01–2006/09.

Time	Constant	FS: Floor space	Age: Age of building	TS: Time to the nearest station	TT: Travel Time to CBD	NU: Number of units	RT: Market reservation time	BD: Bus Dummy	$WT \times BD$	Number of Observations	Adjusted R square
198601	4.402	0.007	-0.143	-0.100	-0.048	-0.011	-0.010	1.333	-0.495	416	0.761
198602	4.508	0.066	-0.144	-0.089	-0.099	-0.010	-0.021	-0.323	0.068	528	0.776
198603	4.464	-0.032	-0.110	-0.070	-0.046	-0.007	-0.022	-0.994	0.480	489	0.823
198604	4.413	0.051	-0.161	-0.106	-0.029	0.006	-0.012	1.160	-0.489	455	0.824
198605	4.669	0.012	-0.155	-0.096	-0.095	-0.002	-0.034	0.722	-0.268	605	0.727
198606	4.343	0.057	-0.133	-0.132	-0.025	0.014	-0.020	-0.912	0.268	446	0.751
198607	3.930	0.083	-0.120	-0.116	-0.054	0.002	-0.046	-0.470	0.168	430	0.785
198608	4.401	0.034	-0.128	-0.113	-0.009	0.021	-0.058	0.078	-0.095	564	0.791
198609	4.526	0.069	-0.146	-0.146	-0.018	0.008	-0.020	-0.308	0.084	394	0.838
198610	4.250	0.043	-0.114	-0.092	-0.025	0.014	-0.024	0.177	-0.122	560	0.826
198611	4.310	-0.001	-0.109	-0.121	0.053	0.010	-0.031	0.036	-0.080	340	0.866
198612	4.822	0.077	-0.154	-0.133	-0.142	-0.006	-0.032	-1.199	0.425	342	0.896
199001	5.831	-0.114	-0.154	-0.084	-0.067	0.022	0.005	-0.198	0.025	857	0.763
199501	4.820	0.090	-0.208	-0.070	-0.048	0.011	0.044	-0.203	0.064	1,109	0.641
200001	4.402	0.071	-0.209	-0.036	-0.035	0.021	0.005	-0.373	0.125	778	0.697
200501	4.548	0.035	-0.208	-0.057	-0.015	0.018	0.009	-0.752	0.294	702	0.757
200601	4.303	0.084	-0.191	-0.093	-0.011	0.023	-0.002	-0.576	0.188	650	0.809
200602	4.484	0.115	-0.201	-0.106	-0.040	-0.002	-0.010	-0.382	0.125	768	0.766
200603	4.584	0.056	-0.201	-0.087	-0.024	0.025	-0.007	-0.670	0.207	1,015	0.785
200604	4.441	0.067	-0.182	-0.091	-0.024	0.012	-0.011	-0.242	0.051	826	0.775
200605	4.583	0.060	-0.186	-0.080	-0.036	0.015	-0.014	-0.224	0.049	966	0.774
200606	4.807	0.033	-0.208	-0.072	-0.022	0.000	-0.018	-0.248	0.039	776	0.775
200607	4.530	0.063	-0.187	-0.092	0.001	0.007	-0.006	-0.039	-0.043	819	0.770
200608	4.742	0.024	-0.198	-0.079	-0.015	0.011	-0.011	0.324	-0.135	901	0.784
200609	4.566	0.020	-0.198	-0.072	0.012	0.018	-0.005	0.089	-0.070	900	0.766

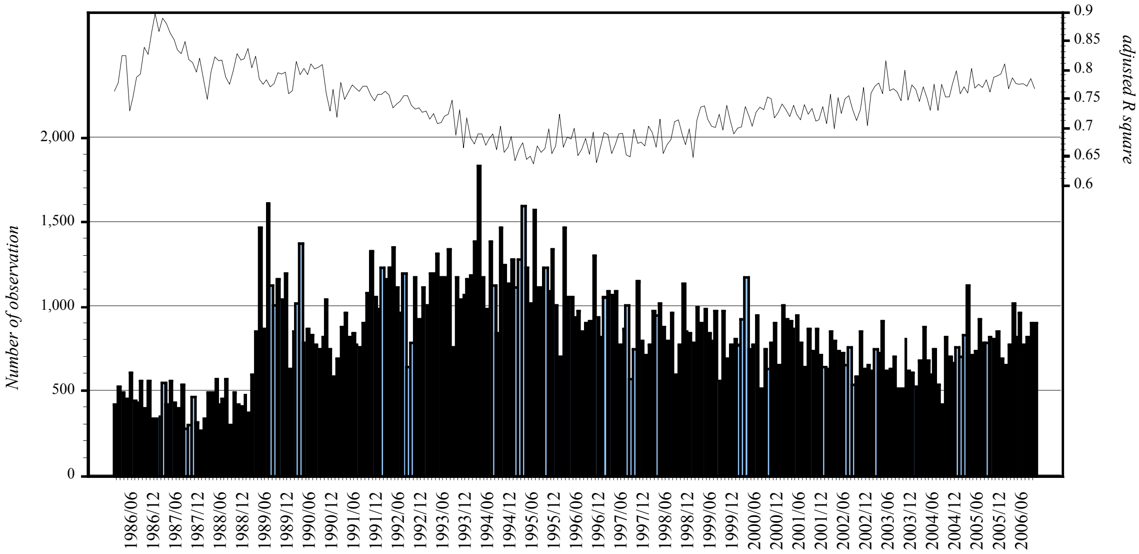

Figure1.Estimation accuracy of the URHM: between 1986/01 and 2006/09.

Table 5. Statistical values of major regression coefficients (URHM).

Principal Independent Variables	***RHI*** **:1986.01 - 2006.09**	***NRHI*** :Summary statistics of estimated parameter			
		Average	Standard deviation	Skewness	Kurtosis
FS :Floor space/square meters	0.013	0.033	0.081	-0.758	-0.627
Age :Age of building	-0.190	-0.185	0.033	0.474	0.110
WT :Distance to nearest station	-0.078	-0.082	0.019	-0.640	0.799
TT :Travel Time to central business district	-0.040	-0.041	0.032	-0.320	0.136
Adjusted-R Square	0.837	0.741	0.054	0.190	-0.379
Number of Samples	211,179	844.720	282.977	0.369	0.123

1986.01 - 2006.09:Monthly ,Number of Mode=250

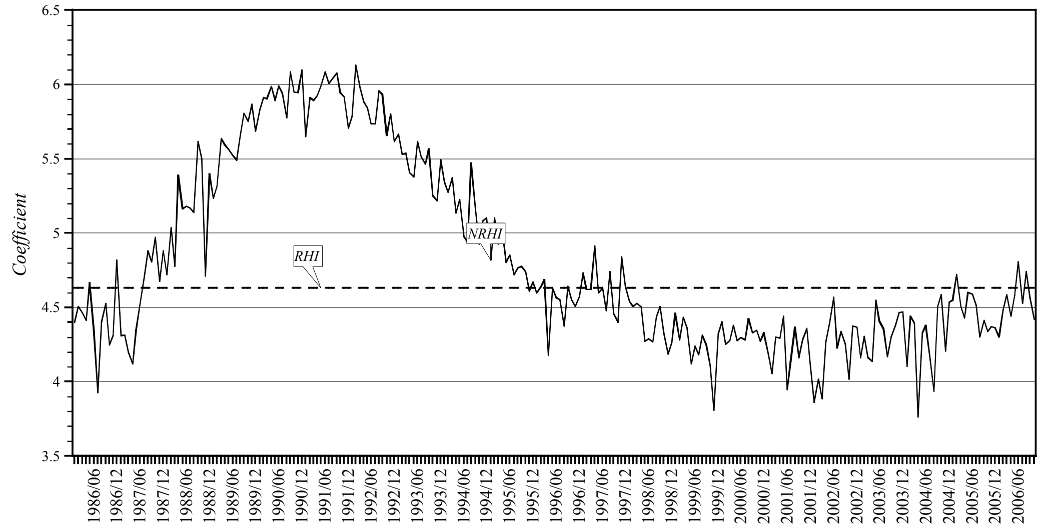

Fig2. Time profile of regression coefficient of the URHM, constant term *cnst*: 1986/01–2006/09.

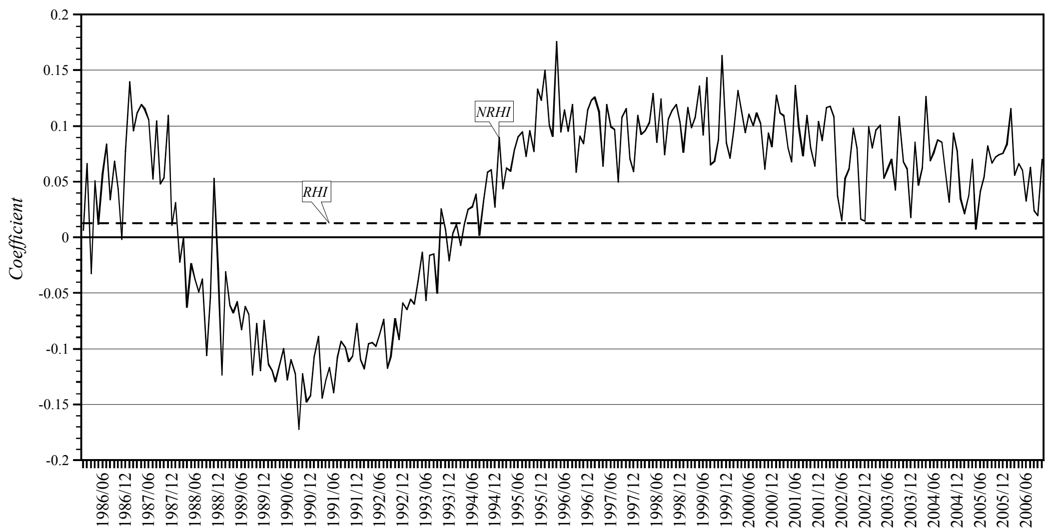

Fig3. Time profile of regression coefficient of the URHM, floor space *FS*: 1986/01–2006/09.

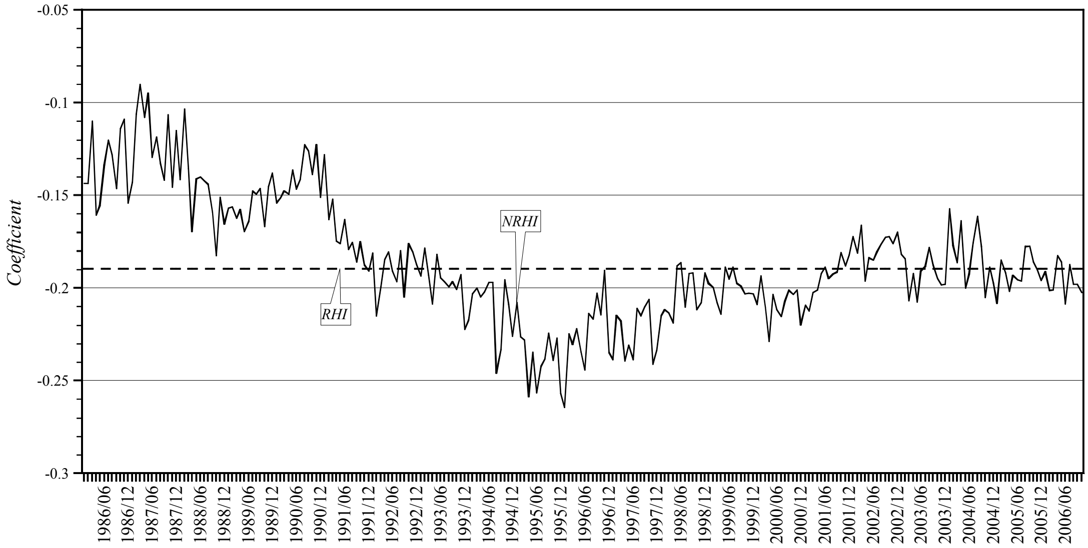

Fig4. Time profile of regression coefficient of the URHM, age of building *Age*: 1986/01–2006/09.

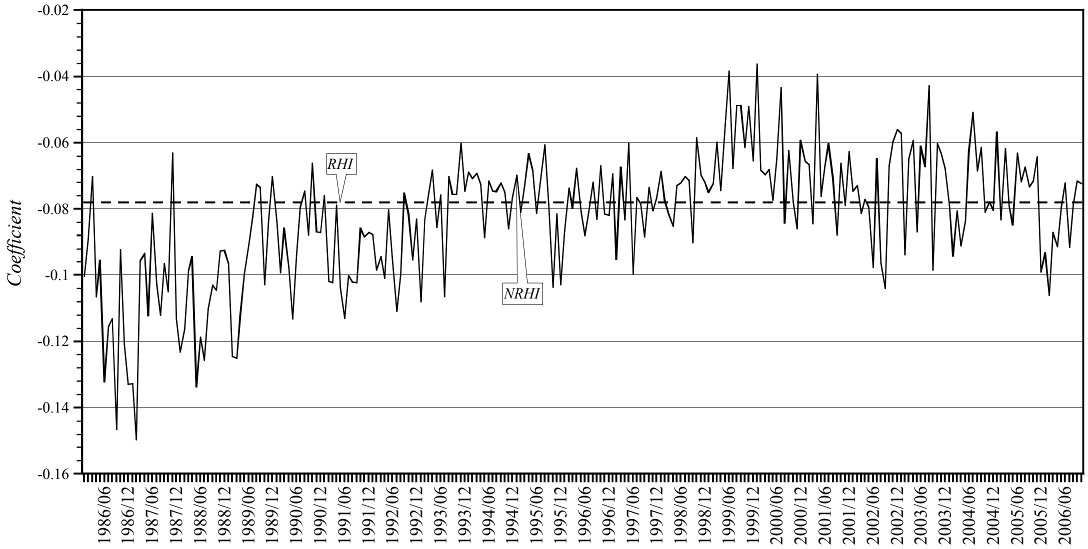

Fig5. Time profile of regression coefficient of the URHM, time to nearest station *TS*: 1986/01–2006/09.

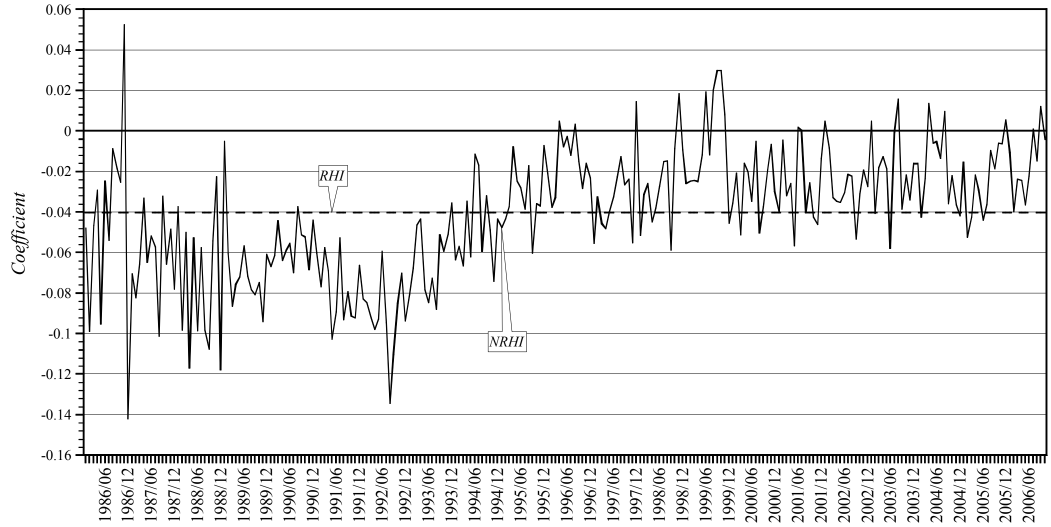

Fig6. Time profile of regression coefficient of the URHM, travel time to CBD *TT*: 1986/01–2006/09.

Thus, structural changes are expected to occur as a background to changes in the regression coefficients. For example, with respect to *FS*, the regression coefficient is estimated as positive in the basic model, while it is estimated as negative from 1987 to the end of 1995. This indicates that the structure apparently differs for this period (1987–1995) compared with other periods.

Meanwhile, large changes occur within short periods, the range of which exceeds that of the long-term changes. For example, when we look at the changes in regression coefficients of the *TS* and the *TT* to the CBD shown in Figs 5 and 6, we can see the tendency of the coefficient line to increase throughout the period. For *Age* (Fig. 4), the trend is a decrease until the mid 1990s, followed by an increase thereafter. While there are such changes observed over all periods, there are major fluctuations within short periods beyond the range of such changes (beyond the difference in the regression coefficient between the beginning and ending periods). These changes in the short periods cannot be considered to be structural changes.

On the basis of these findings, we consider that part of the change in the regression coefficients over time is caused by structural changes. It is also expected that some bias in the observation data group in each period may have a certain effect.

4.3.Estimation of OPHI

4.3.1.Estimation of model and changes in price structure

Next, an OPHM is estimated on the basis of the procedure described in Section2.2. In the OPHM, to absorb changes during a short period, the presence of which became apparent in the estimation of the URHM, data over a certain period are pooled for the estimation. Using such an estimation method, we expect to absorb changes during a short period. However, upon estimation of the OPHM, setting an overlapped estimation period length (τ) is the key.

In the estimation of the URHI, it has been clarified that there is a seasonal characteristic of changes in the thickness (more precisely, the number of transactions) of the housing market, and that the number of transactions increases at the end of the fiscal year between January and March, and the number decreases in July and August. When we attempt to absorb such a seasonally fluctuating characteristic in the market, the estimation period should be set to exceed one year. When the period is set to be longer, parameters are expected to be stable. However, it becomes difficult to accurately represent changes in the market structure. In this study, on the basis of such assumptions, we examine the effects of varying τ between 12 months and 36 months on the price index and on the regression coefficients that represent the price-forming structure of the major variables.

First, estimation results using τ of 12 months are shown in Table 6 and Fig. 7. The coefficient of determination decreases from 1986 through 1995, similar to that in the URHM; then it increases from the second half of 1996. On the whole, the coefficient of determination maintains an average of approximately 0.75, indicating a good result similar to that of the URHM.

Next, we focused on the regression coefficients of the estimated model. Table 7 shows descriptive statistical values of the regression coefficients for 238 periods. Figs 8–12 show temporal changes in the regression coefficients.

When we look at the changes in the regression coefficients over time, the wide fluctuations observed in the URHM are eliminated and smooth changes are shown, making it easier to clarify overall tendencies. The absolute values of the regression coefficients of the *TS* and *TT* to the CBD tend to decrease with time. In other words, elasticity in terms of distance is decreasing. Regarding *Age*, quasi-periodic changes in the regression coefficients are observed. The regression coefficient of *FS* is estimated to be negative from 1989 through 1995, then positive thereafter. It showed a stable value at around 0.1 for several years after 1996, but since then, the elasticity has been gradually decreasing. In summary, we can see that consumer preference is shifting its emphasis from the location to the amount of *FS* and *Age* in recent years.

The coefficients of variation were obtained as follows (Table 7): *FS* = 2.424 (2.428 in the URHM), *Age* = −0.163 (−0.179 in the URHM), *TS* = −0.178 (−0.232 in the URHM), *TT* to CBD = −0.554 (−0.779 in the URHM). On the whole, the values are smaller than those in the URHM. However, the variations of the coefficients of *FS* and *Age* are not significantly different from those of the URHM, indicating that the regression coefficients changed not because of temporal changes caused by sample bias, but because of large structural changes. The coefficients of variance of *TS* and *TT* to the CBD were significantly lower than those in the URHM. This is speculated to be because of large temporal changes in the regression coefficients caused by sample bias existing in a unit of time.

Next, in Figs 8–12, we observed changes in the regression coefficients when τ was varied between 12 months and 36 months. Compared with the estimation results of the URHM, with regard to the temporal changes in the regression coefficients of each variable, we can see the presence of a time lag in the changes in regression coefficients as τ becomes longer. This tendency is commonly observed in all the variables. Such a time lag is also expected to affect the price indices.

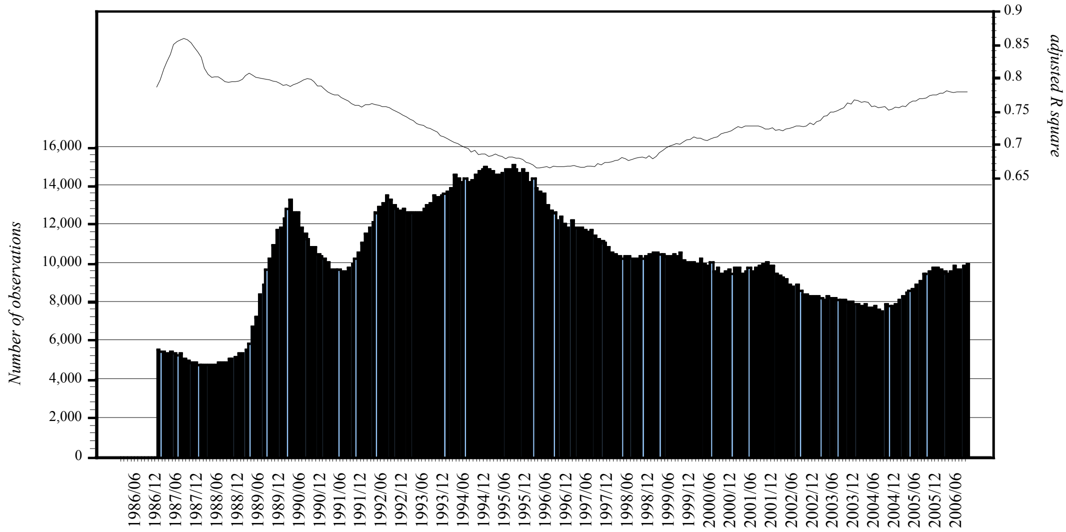

Fig7. Estimation accuracy of the OPHM: between 1986/01 and 2006/09.

Table6. Estimation results of the OPHM ($\tau = 12$): 1986/01–2006/09.

Time	Constant	*FS*: Floor space	*Age*: Age of building	*TS*: Time to the nearest station	*TT*: Travel Time to CBD	*NU*: Number of units	*RT*: Market reservation time	*BD*: Bus Dummy	*WT* × *BD*	Number of Observations	Adjusted R square
198612	4.232	0.041	-0.129	-0.108	-0.046	0.002	-0.028	-0.156	0.035	5,497	0.785
198701	4.176	0.055	-0.129	-0.112	-0.047	0.003	-0.028	-0.106	0.010	5,425	0.796
198702	4.133	0.058	-0.126	-0.117	-0.043	0.007	-0.025	-0.075	-0.009	5,446	0.812
198703	4.067	0.068	-0.126	-0.120	-0.045	0.008	-0.022	-0.023	-0.034	5,383	0.824
198704	4.038	0.075	-0.123	-0.122	-0.050	0.011	-0.021	-0.088	-0.006	5,492	0.835
198705	3.994	0.085	-0.120	-0.124	-0.048	0.014	-0.019	-0.133	-0.005	5,316	0.849
198706	4.015	0.089	-0.122	-0.119	-0.050	0.013	-0.021	-0.095	-0.019	5,268	0.854
198707	4.053	0.087	-0.121	-0.118	-0.050	0.013	-0.021	-0.053	-0.037	5,372	0.856
198708	4.104	0.089	-0.120	-0.115	-0.057	0.012	-0.018	-0.049	-0.041	5,083	0.859
198709	4.139	0.091	-0.120	-0.111	-0.059	0.012	-0.020	-0.084	-0.031	4,986	0.857
198710	4.183	0.091	-0.119	-0.111	-0.059	0.012	-0.020	-0.162	0.002	4,888	0.852
198711	4.293	0.093	-0.123	-0.108	-0.066	0.010	-0.016	-0.231	0.024	4,863	0.846
198712	4.315	0.092	-0.121	-0.109	-0.066	0.012	-0.009	-0.257	0.033	4,792	0.839
199001	5.522	-0.083	-0.154	-0.092	-0.074	0.020	0.007	-0.275	0.049	12,360	0.788
199501	5.191	0.030	-0.209	-0.074	-0.047	0.019	0.030	-0.360	0.105	14,903	0.681
200001	4.244	0.105	-0.201	-0.059	-0.008	0.021	0.008	-0.307	0.084	10,033	0.710
200501	4.327	0.069	-0.185	-0.076	-0.020	0.023	0.000	-0.172	0.027	8,131	0.755
200601	4.454	0.057	-0.189	-0.075	-0.026	0.018	-0.006	-0.404	0.117	9,684	0.776
200602	4.443	0.062	-0.191	-0.076	-0.025	0.017	-0.006	-0.461	0.141	9,622	0.777
200603	4.450	0.063	-0.191	-0.079	-0.024	0.017	-0.006	-0.544	0.172	9,506	0.780
200604	4.439	0.064	-0.190	-0.080	-0.023	0.017	-0.007	-0.486	0.149	9,617	0.778
200605	4.438	0.069	-0.189	-0.080	-0.024	0.016	-0.007	-0.486	0.148	9,844	0.777
200606	4.446	0.069	-0.189	-0.082	-0.022	0.014	-0.008	-0.444	0.130	9,699	0.778
200607	4.449	0.070	-0.189	-0.083	-0.018	0.014	-0.008	-0.409	0.118	9,726	0.777
200608	4.485	0.064	-0.190	-0.084	-0.018	0.014	-0.007	-0.347	0.098	9,837	0.778
200609	4.494	0.060	-0.192	-0.083	-0.015	0.015	-0.008	-0.249	0.061	9,920	0.778

Table7. Statistical values of major regression coefficients ($\tau = 12$).

Principal Independent Variables	***RHI*** **:1986.01 - 2006.09**	***OPHM*** :Summary statistics of estimated parameter			
		Average	Standard deviation	Skewness	Kurtosis
FS :Floor space/square meters	0.013	0.033	0.079	-0.880	-0.770
Age :Age of building	-0.190	-0.185	0.030	0.545	-0.392
TS :Time to nearest station	-0.078	-0.082	0.015	-0.984	0.214
TT :Travel Time to central business district	-0.040	-0.042	0.023	-0.518	-1.084
Adjusted-R Square	0.837	0.738	0.049	0.237	-0.745
Number of Samples	211,179	10,178.252	2,709.339	-0.198	-0.467

1986.12 - 2006.09:Monthly ,Number of Mode=238

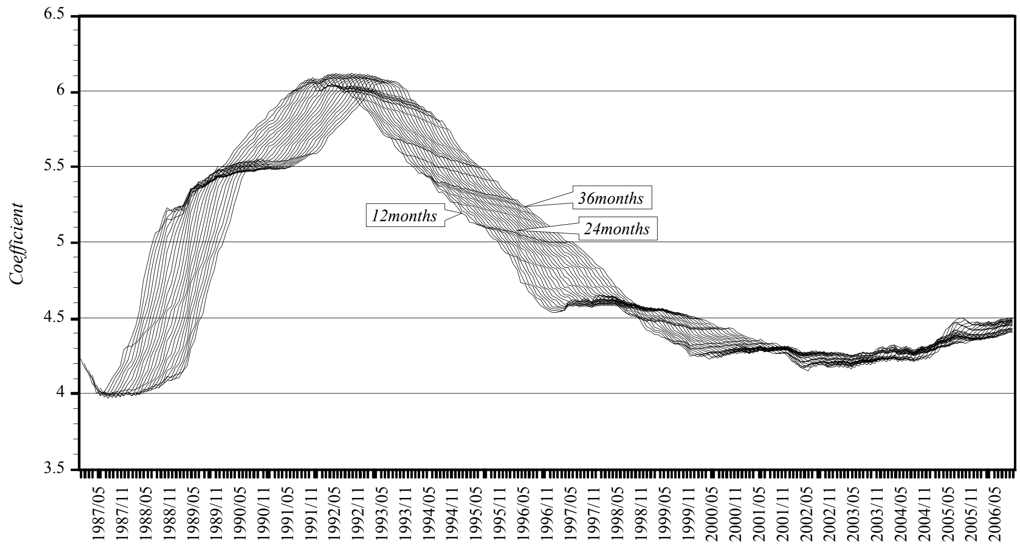

Fig. 8. Time profile of regression coefficient of the OPHM, constant term *cnst*: 1986/01–2006/09.

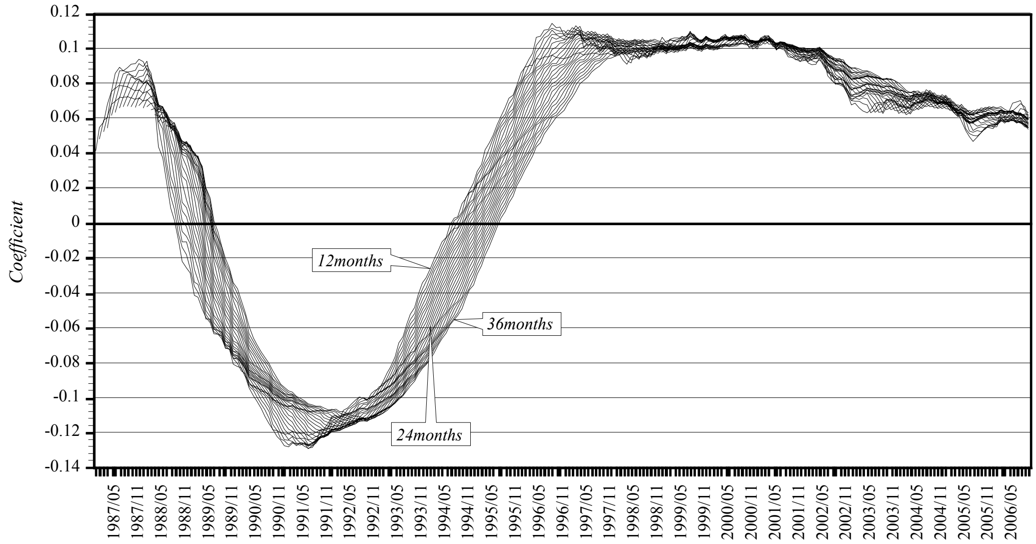

Fig9. Time profile of regression coefficient of the OPHM, floor space *FS*: 1986/01–2006/09.

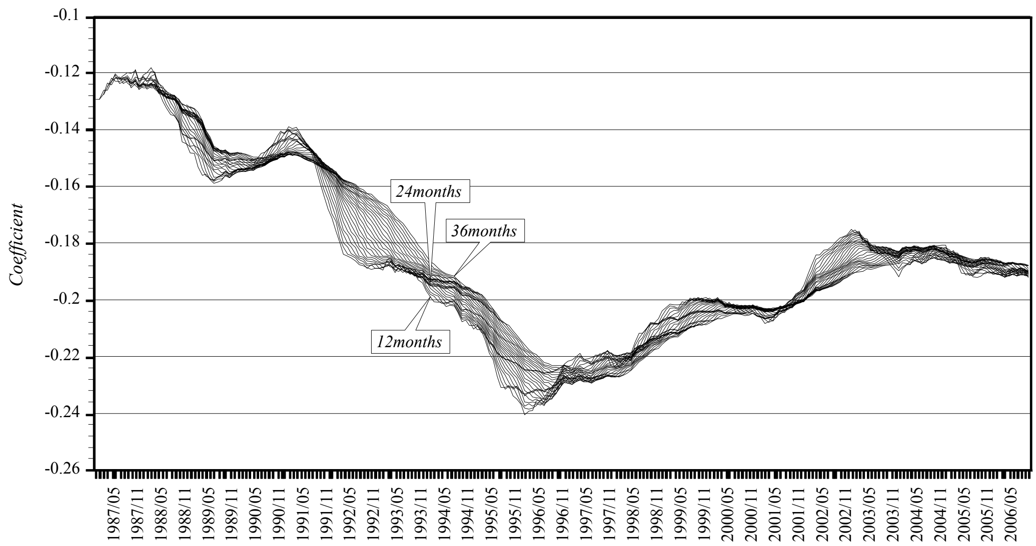

Fig10. Time profile of regression coefficient of the OPHM, age of building *Age*: 1986/01–2006/09.

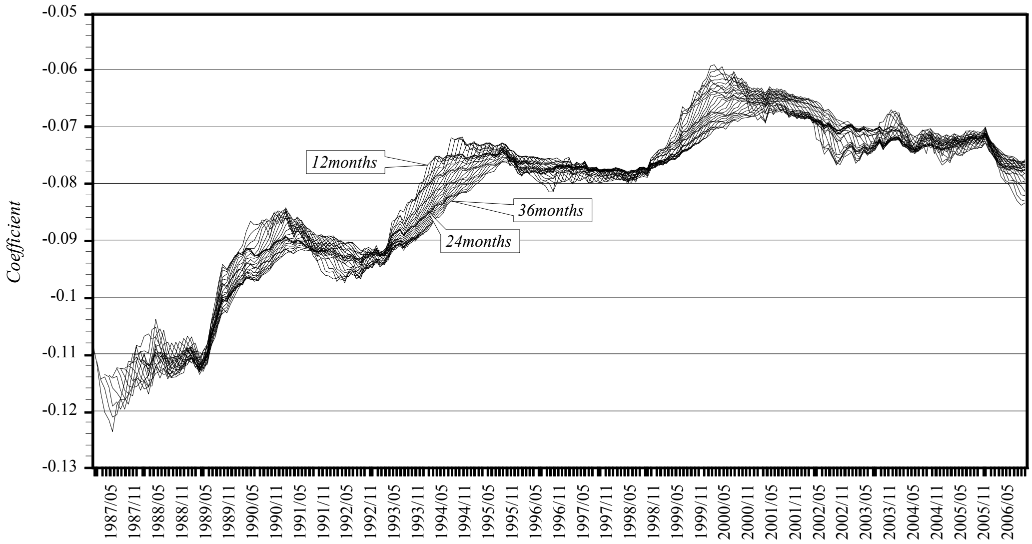

Fig11. Time profile of regression coefficient of the OPHM, time to nearest station *TS*: 1986/01–2006/09.

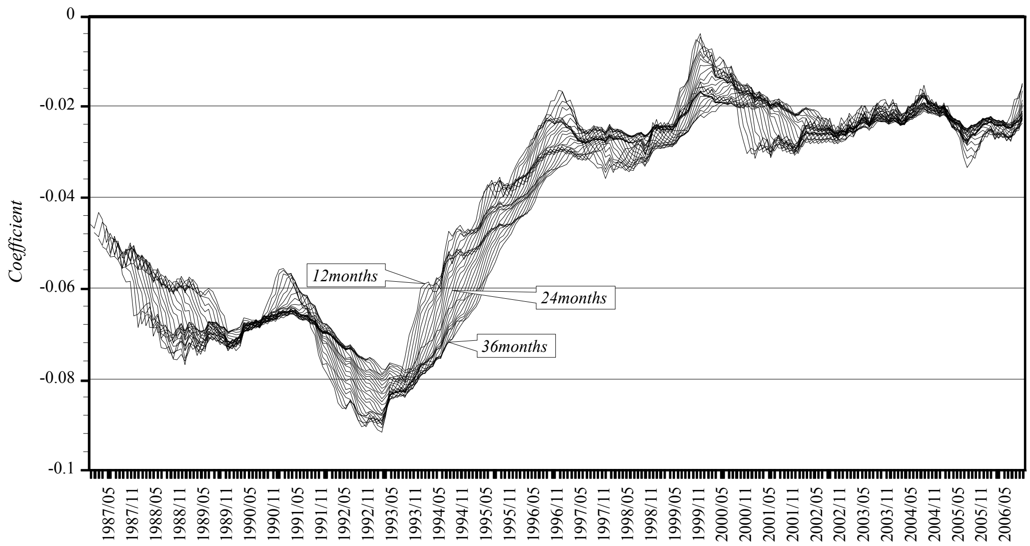

Fig12. Time profile of regression coefficient of the OPHM, travel time to CBD *TT*: 1986/01–2006/09.

4.3.2.Evaluation of indices by different τ

We set τ in order to avoid the effects of specific bias in the observation data when the data were divided monthly; the bias, if present, is reflected in the regression coefficients and may generate effects that are difficult to differentiate from structural changes. The background of the bias is not clear at this stage of the study, but from the monthly samples, seasonal changes in the timing of when properties are put on the market were confirmed. Therefore, if we wish to avoid the seasonal fluctuation characteristic, at least 12-month periods are necessary. Namely, the data division should accommodate the bias resulting from seasons with or without a large movement of people. However, when τ was extended, the presence of a time lag relative to changes in the market was also observed.

Therefore, we varied τ between 12 months and 36 months, and observed the changes in the price indices, as shown in Fig. 13.

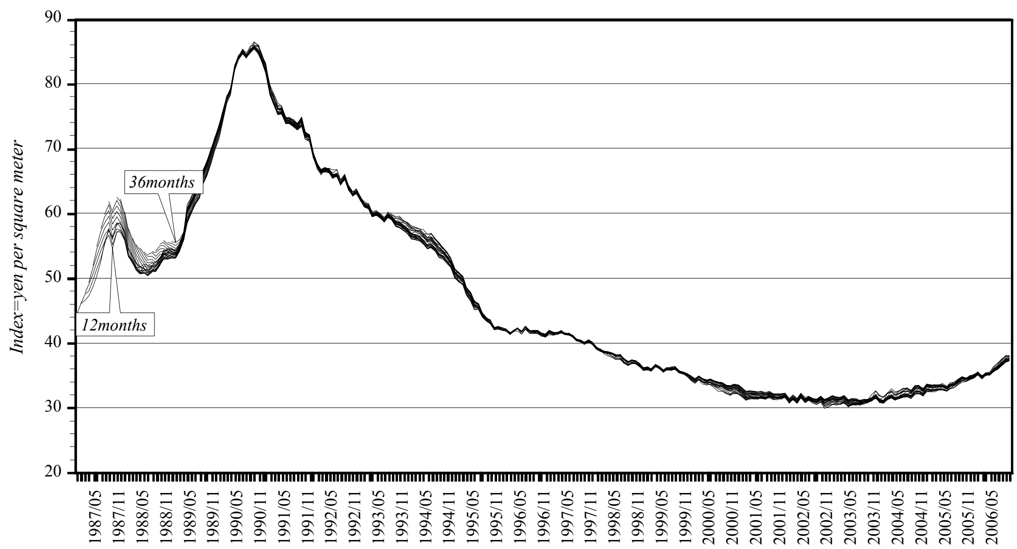

Fig13. Time profile of OPHI: 1986/01–2006/09.

We can see that the time lag in the regression coefficients also affects the price indices. When we compare changes in the price indices (Fig. 13) with changes in the constant term (Fig. 8), we can see that a time lag is generated in the coefficient of the constant term (Fig. 8) as the connection period length increases. In contrast, for the price index (Fig. 13), as the connection period length increases, the increase in price occurs earlier. In other words, as the connection period length increases, while the price change excluding the effects of regression coefficients is delayed, the increase in the price index occurs at an earlier period than the actual time of the price increase because the effect of the price increase in the future is incorporated in advance. This effect is particularly strong in the bubble-economy years of 1986 and after, when the price increase was particularly rapid.

These findings indicate that τ should be short for the estimation of the price index. Then, because it is necessary to overlap at least 12 months to exclude the seasonal change in the number of transactions, the optimal τ is determined to be 12 months.

4.4.Comparison between RHI/ URHI and OPHI

In the analysis described in the previous section, the OPHI with $\tau = 12$ months was shown to be the most accurate in representing the market trends. Here, we compare the RHI, the URHI, and the OPHI for $\tau = 12$ months (Fig. 14).

When these indices are compared, large fluctuations in the URHI are noted (Fig. 14). The magnitude of these fluctuations seems to be different from our actual experience of price changes, because we did not experience any large increase or decrease in the prices of condominiums with specific qualities during the periods when the indices showed a large increase or decrease. In particular, because the fluidity of the condominium market is low, instantaneous changes in the price are not expected.

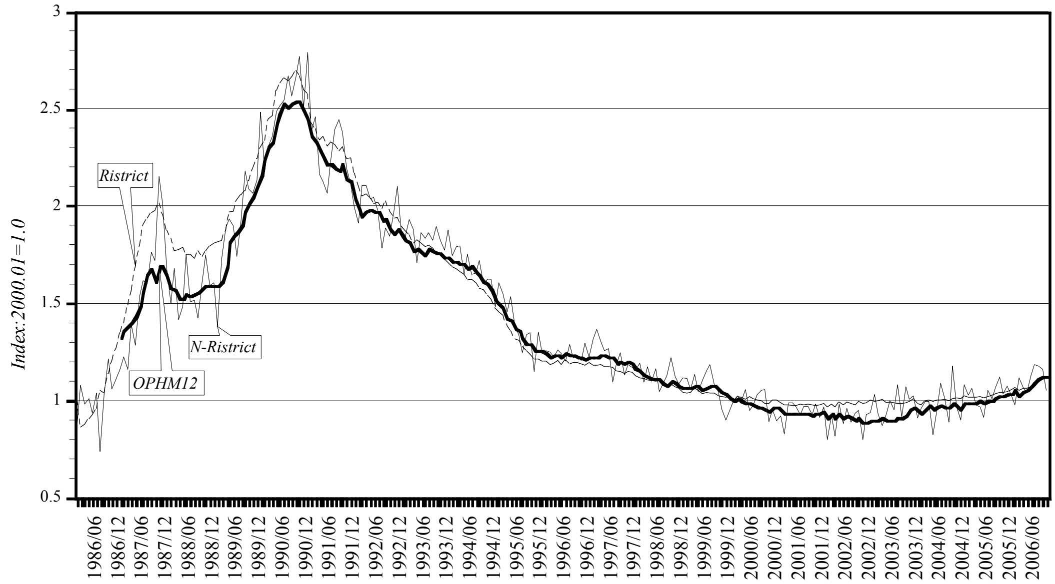

Fig14. Comparison of indices between RHI,URHI and OPHI(τ =12): 1986/01–2006/10.

The magnitude of fluctuation in an index is not a priori a point of evaluation of the quality of the index. However, such excessive fluctuations cannot be justified. In particular, while large changes were observed in the regression coefficient, it is difficult to expect that consumers in the market randomly and significantly changed their preferences; therefore, we speculate that the regression coefficient largely changed because of bias in the data, and that, as a result, fluctuations in the price index occurred.

When we compare the RHI with the OPHI, there are large differences between the two from 1986 through 1990. When we further compare the URHI with the above two indices, the URHI showed random fluctuations centered around the OPHI. While the URHI enables us to represent market changes most sensitively, it fluctuates considerably because of the large bias in samples. Considering these characteristics, rather than the RHI, the URHI can represent changes in the market structure more appropriately and is expected to be more accurate.

The results of the above analysis demonstrated that among the RHI, URHI, and OPHI (τ = 12 months), the OPHI (τ = 12 months) is superior, because the RHM is not able to accurately estimate the market trends under the circumstance of large changes in the price level and the price structure during the years from 1986 through 1990, and because the URHI fluctuates considerably because of the seasonal fluctuation of transactions as well as other biases.

5.Conclusions

In the series of analyses described above, we focused on temporal changes in the house (condominium) price structure, and examined the effects of structural change and seasonal sample bias on price indices. As a result of the analyses, we clarified the following.

First, when using the model of URHI, which takes into account changes in the market structure, the regression coefficients widely fluctuate in each period or every couple of periods. Although this fluctuation is observed within short periods, a specific trend is observed in the long term. The fluctuation is speculated to be because of not only the occurrence of structural changes in the condominium market, but also the occurrence of bias in the transaction samples in each period. More precisely, we find that there is a seasonal fluctuation in the condominium market, which leads to active or sluggish transactions depending on the time of year, and that such changes affect the regression coefficients.

To eliminate the bias existing in transaction samples in each period, which are mainly exhibited as the seasonal fluctuation, and to eliminate changes in the market structure, we proposed OPHM. Here, to eliminate the bias in the number of transactions and in samples affected by the seasonal change in the number of transactions, we set τ from 12 months to 36 months for the estimation. The results depicting the changes in major regression coefficients over time showed gradual changes instead of the wide fluctuations observed in the URHM, thus, it became easier to understand the trend.

Temporal changes in the regression coefficients revealed by OPHM showed that the absolute values of the regression coefficients of *TS* and *TT* to the CBD have tended to gradually decrease in recent years, indicating that the elasticity in terms of distance has become smaller. There were slight periodic changes in the regression coefficient of *Age*, and the regression coefficient of *FS* was estimated to be negative from 1989 through 1995, then positive thereafter. These results demonstrate that consumer preference is moving towards the amount of *FS* and *Age*, rather than location.

When τ between 12 months and 36 months were compared, the existence of a time lag in the regression coefficients was observed. This time lag in the regression coefficients from 1986 during the bubble-economy years when prices were rapidly increasing also resulted in a large time lag in the price index. From these findings, τ should be short. In addition, considering the aim of resolving the bias because of seasonal changes in the number of transactions, the most suitable τ was determined to be 12 months.

The comparison among RHI, URHI and OPHI (τ = 12 months) showed that there are large fluctuations in URHI, and that the magnitude of the fluctuations differs significantly from that in actual prices.

The comparison between URHI and OPHI showed the presence of large differences between the two indices from 1986 through 1990. When URHI is added to this comparison, we can see random changes in the URHI centered around the values of the OPHI. From the results of the comparison of these three price indices, it was speculated that, because there were large changes in the price structure from 1986 through 1990, RHI was unable to accommodate such changes, so that large differences between URHI and URHI/OPHI (τ = 12 months) were observed.

The results of the above series of analyses indicated the superiority of the estimation by OPHI (τ = 12 months) in the secondhand condominium market in Japan, when structural changes in the market are to be accommodated.

Appendix. Statistical nature of RHI and URHI

Characteristics of RHI and URHI

The characteristics of price indices generated by the two methods are summarized. Some of the characteristics, which are basic principles of regression analysis, will also be noted in relation to the hedonic housing price index.

The principles set in this paper are modified as shown below, using different notations. First, we attempt to obtain a price index using only two periods, i.e., period 0 and period 1, with period 0 as the reference point. The price data in each period are expressed as $y_0(n_0 \times 1)$, $y_1(n_1 \times 1)$, and the explanatory variable data are expressed as $X_0(n_0 \times K)$, $X_1(n_1 \times K)$. Here, the notation of $(n \times m)$ represents the size of the matrix, i.e., number of rows × number of columns. n_0 and n_1 are the number of items of observation data in each period. k is the number of explanatory variables (including the constant term). The RHM is given below, using the pooled observation data for periods 0 and 1 as follows.

$$\tilde{\boldsymbol{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{u}} \quad , \tag{28}$$

where:

$$\tilde{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \end{pmatrix}, \tilde{X} = \begin{pmatrix} \mathbf{0} & X_0 \\ \mathbf{1} & X_1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} a_1 \\ \mathbf{b} \end{pmatrix}, \tilde{\mathbf{u}} = \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \end{pmatrix} \quad .$$

$\boldsymbol{0}$ is a column vector in which all elements are zero.

$\boldsymbol{1}$ is a column vector in which all elements are one.

The first columns of both X_0 and X_1 are vector $\boldsymbol{1}$ because of the constant terms.

The first column $(\boldsymbol{0}', \boldsymbol{1}')$ of $\tilde{X}$ is a time dummy variable of period 1.

a_1 is the regression coefficient of the time dummy variable of period 1.

$\boldsymbol{b}(K \times 1)$ is a regression coefficient vector corresponding to the explanatory variables (including the constant term) from which the time dummy variable is excluded.

$\boldsymbol{\beta}(\overline{K+1} \times 1)$ is a regression coefficient vector comprising a_1 and $\boldsymbol{b}(K \times 1)$.

Both $\boldsymbol{u}_0(n_0 \times 1)$ and $\boldsymbol{u}_1(n_1 \times 1)$ are random disturbance vectors.

In contrast, the URHM is given by the equations below, using the observation data for each period.

$$\boldsymbol{y}_0 = X_0 \boldsymbol{\beta}_0 + \boldsymbol{u}_0 \tag{29}$$

$$\boldsymbol{y}_1 = X_1 \boldsymbol{\beta}_1 + \boldsymbol{u}_1 \tag{30}$$

Characteristic 1: The sum of the errors between the observed value and the estimated value of the price for each period is zero.

In the case of the RHM, from $\hat{\boldsymbol{\beta}} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{\boldsymbol{y}}$, $\widetilde{\mathbf{y}}$ is estimated as $\hat{\boldsymbol{y}} = \widetilde{X}\hat{\boldsymbol{\beta}} = \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{\boldsymbol{y}}$ and the error is given as $\hat{\boldsymbol{u}} = \widetilde{\boldsymbol{y}} - \hat{\boldsymbol{y}} = (I - \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}')\widetilde{\boldsymbol{y}}$. By multiplying both sides of this equation by $\widetilde{X}'$ from the left, we derive $\widetilde{X}'\hat{\boldsymbol{u}} = \boldsymbol{0}$. The first column vector of the data matrix $\widetilde{X}$ is the time dummy variable for period 1 and the second column vector is 1 (because it corresponds to the constant term). Then, $\widetilde{X}'\hat{\boldsymbol{u}} = \boldsymbol{0}$ indicates that the sum of the estimation errors for period 0 = 0 and the sum of the estimation errors for period 1 = 0. The same result can be derived for the URHM .

Characteristic 2: The estimated equation $\hat{\boldsymbol{y}} = \widetilde{X}\hat{\boldsymbol{\beta}}$ of the RHM passes through the center of gravity of the observation data of each period.

From the estimated equation $\hat{\boldsymbol{y}} = \widetilde{X}\hat{\boldsymbol{\beta}}$, we obtain:

$$\hat{\boldsymbol{y}}_0 = X_0 \hat{\boldsymbol{b}}, \quad \hat{\boldsymbol{u}}_0 = \widetilde{\boldsymbol{y}}_0 - \hat{\boldsymbol{y}}_0 \rightarrow \widetilde{\boldsymbol{y}}_0 - \hat{\boldsymbol{u}}_0 = X_0 \hat{\boldsymbol{b}} \tag{31}$$

and

$$\hat{\boldsymbol{y}}_1 = \hat{a}_1 + X_1\hat{\boldsymbol{b}},\ \hat{\boldsymbol{u}}_1 = \tilde{\boldsymbol{y}}_1 - \hat{\boldsymbol{y}}_1 \rightarrow \tilde{\boldsymbol{y}}_1 - \hat{\boldsymbol{u}}_1 = \hat{a}_1 + X_1\hat{\boldsymbol{b}}. \tag{32}$$

If both sides are multiplied by $\boldsymbol{1}'$ (vector with all the elements equal to 1) and the results are divided by the number of items of observation data of each period, with consideration given to characteristic 1, we then obtain:

$$\bar{y}_0 = \bar{\boldsymbol{x}}_0'\hat{\boldsymbol{b}} \tag{33}$$

and

$$\bar{y}_1 = \hat{a}_1 + \bar{\boldsymbol{x}}_1'\hat{\boldsymbol{b}}. \tag{34}$$

We find $\bar{y}_0$ and $\bar{y}_1$ to be the average prices and $\bar{\boldsymbol{x}}_0$ and $\bar{\boldsymbol{x}}_1$ to be the averages of the explanatory variables in the two periods. Similarly, in the URHM , we obtain:

$$\bar{y}_0 = \bar{\boldsymbol{x}}_0'\hat{\boldsymbol{\beta}}_0 \tag{35}$$

and

$$\bar{y}_1 = \bar{\boldsymbol{x}}_1'\hat{\boldsymbol{\beta}}_1. \tag{36}$$

Characteristic 3: If the averages of the explanatory variables of period 0 and period 1 are equal ($\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}_1$), then the RHI and the URHI of the houses having this average value of quality are equal.

When $\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}_1 = \bar{\boldsymbol{x}}$, the estimated house prices with the quality value of $\bar{\boldsymbol{x}}$ are as below, from characteristic 2.

$$\bar{y}_0 = \bar{\boldsymbol{x}}'\hat{\boldsymbol{b}} = \bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}}_0 \tag{37}$$

$$\bar{y}_1 = \hat{a}_1 + \bar{\boldsymbol{x}}'\hat{\boldsymbol{b}} = \bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}}_1 \tag{38}$$

This means that the estimated house prices using both models (RHM and URHM) for houses with the quality value of $\bar{\boldsymbol{x}}$ are the average house prices in each period. Therefore, characteristic 3 holds.

The relationship with the URHI

In addition to the RHM of equation (1) and the URHM based on equations (9) and (10), we add another RHM without time dummy variables.

$$\tilde{\boldsymbol{y}} = X\boldsymbol{\beta}_* + \boldsymbol{u}_* \tag{39}$$

Here,

$$\tilde{\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \end{pmatrix}, \quad X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}, \quad \boldsymbol{u}_* = \begin{pmatrix} \boldsymbol{u}_{*0} \\ \boldsymbol{u}_{*1} \end{pmatrix}$$

holds. $\boldsymbol{\beta}_*$ represents the regression coefficient vector, which contains a constant term but does not contain time dummy variables.

Characteristic 4: If the variance of the random disturbance term of each period is the same, the regression coefficient $\hat{\boldsymbol{\beta}}_*$ of the RHM without the time dummy variable is the weighted average of the regression coefficients $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ of the URHM using the inverse matrix of the respective variance–covariance matrix of the regression coefficient.

The estimated value of $\boldsymbol{\beta}_*$ of the model based on equation (10) is as follows.

$$\begin{aligned} \hat{\boldsymbol{\beta}}_* &= (X'X)^{-1}X'\tilde{\boldsymbol{y}} \\ &= (X_0'X_0 + X_1'X_1)^{-1}(X_0'\boldsymbol{y}_0 + X_1'\boldsymbol{y}_1) \end{aligned} \tag{40}$$

Meanwhile, the estimated values of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ of the URHM based on equation (7) and (8) are as follows.

$$\hat{\boldsymbol{\beta}}_0 = (X_0'X_0)^{-1}X_0'\boldsymbol{y}_0 \tag{41}$$

$$\hat{\boldsymbol{\beta}}_1 = (X_1{}'X_1)^{-1}X_1{}'\boldsymbol{y}_1 \tag{42}$$

Assuming that $\boldsymbol{u}_0 \sim N(\boldsymbol{0}, \sigma^2 I_0)$ and $\boldsymbol{u}_1 \sim N(\boldsymbol{0}, \sigma^2 I_1)$, namely, the variance of the random disturbance term in each period is the same, we obtain:

$$Var(\hat{\boldsymbol{\beta}}_0) = \sigma^2 (X_0{}'X_0)^{-1} \quad (= V_0,) \tag{43}$$

$$Var(\hat{\boldsymbol{\beta}}_1) = \sigma^2 (X_1{}'X_1)^{-1} \quad (= V_1) \ . \tag{44}$$

Equation (28) can then be rewritten using equations (29), (30), (31) and (32) as follows.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_* &= (\sigma^2 V_0^{-1} + \sigma^2 V_1^{-1})^{-1}(\sigma^2 V_0^{-1}\hat{\boldsymbol{\beta}}_0 + \sigma^2 V_1^{-1}\hat{\boldsymbol{\beta}}_1) \\ &= (V_0^{-1} + V_1^{-1})^{-1}(V_0^{-1}\hat{\boldsymbol{\beta}}_0 + V_1^{-1}\hat{\boldsymbol{\beta}}_1)\end{aligned} \tag{45}$$

Namely, the regression coefficient $\hat{\boldsymbol{\beta}}_*$ of the RHM is obtained by calculating the weighted average of the regression coefficients $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ of the URHM using the inverse matrix of the respective variance–covariance matrix of the regression coefficient.

Actually, also in the case of the RHM represented by equation (26), which includes time dummy variables, the regression coefficients other than the constant term and time dummy variables are the weighted averages of the corresponding regression coefficients of the URHM using the inverse matrix of the respective variance–covariance matrix. This is simply because the RHM represented by equation (26) is obtained by shifting the regression plane of the RHM without the time dummy variables, represented by equation (28), parallel as it passes through the center of gravity of the observation data for each period (Characteristic 2). Consequently, the regression coefficients of equations (26) and (28) are the same except for the constant term and time dummy variables.

Characteristic 5: The RHI and the URHI are identical for those houses whose property characteristics are defined by calculating the weighted averages of the explanatory variables for each period using the variance–covariance matrix of the regression coefficients $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ of the URHM.

As is evident from the method of deriving structurally restricted indices described in Section 2.1, the RHI for period 1 generated using the RHM based on equation (26) is represented as a time dummy variable $\hat{a}_1$. Let us consider a case where the URHI and restricted price index have the same value. Namely, the URHI of houses with quality m for period 1 can be described as follows.

$$\hat{y}_1 - \hat{y}_0 = \boldsymbol{m}'\hat{\boldsymbol{\beta}}_1 - \boldsymbol{m}'\hat{\boldsymbol{\beta}}_0 = \boldsymbol{m}'(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) \tag{46}$$

We attempt to solve this equation to find the value of m for which equation (46) equals $\hat{a}_1$. This helps us understand under which conditions the two indices, i.e., the RHI and URHI, show similar movements.

Provided that $\bar{y}_0$ and $\bar{y}_1$ are the mean prices, and $\bar{\boldsymbol{x}}_0$ and $\bar{\boldsymbol{x}}_1$ are the averages of the explanatory variables in each period, the RHI $\hat{a}_1$ is expressed as follows.

$$\begin{aligned} \hat{a}_1 &= (\bar{y}_1 - \bar{\boldsymbol{x}}_1'\hat{\boldsymbol{\beta}}_*) - (\bar{y}_0 - \bar{\boldsymbol{x}}_0'\hat{\boldsymbol{\beta}}_*) \\ &= (\bar{y}_1 - \bar{y}_0) - (\bar{\boldsymbol{x}}_1' - \bar{\boldsymbol{x}}_0')\hat{\boldsymbol{\beta}}_* \end{aligned} \tag{47}$$

This equation can be transformed by $\hat{\boldsymbol{\beta}}_* = (V_0^{-1} + V_1^{-1})^{-1}(V_0^{-1}\hat{\boldsymbol{\beta}}_0 + V_1^{-1}\hat{\boldsymbol{\beta}}_1)$, which is derived from the analysis of equation (27), as follows.

$$\hat{a}_1 = (\bar{\boldsymbol{x}}_0'W_0 + \bar{\boldsymbol{x}}_1'W_1)(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) \tag{48}$$

We know that:

$$W_0 = V_0(V_0 + V_1)^{-1} \tag{49}$$

$$W_1 = V_1(V_0 + V_1)^{-1} \tag{50}$$

hold. Therefore, we can obtain the value of *m* for which $\boldsymbol{m}'(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) = \hat{a}_1$ as follows.

$$\boldsymbol{m}' = \bar{\boldsymbol{x}}_0{}' W_0 + \bar{\boldsymbol{x}}_1{}' W_1 \tag{51}$$

This shows that *m* is obtained by a weighted averaging of the averages of the explanatory variables for each period using the variance–covariance matrix of the regression coefficients $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ of the URHM . This once more demonstrates the points described in Characteristic 3, because $\boldsymbol{m} = \bar{\boldsymbol{x}}$ holds when $\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}_1 = \bar{\boldsymbol{x}}$.

References

Bourassa, ST, Hoesli M, Sun J. A simple alternative house price index method. Journal of Housing Economics 2006;15; 80–97.

Case B, Quigley JM. The dynamics of real estate prices. Review of Economics and Statistics 1991;22; 50–58.

Case KE, Shiller RJ. Prices of single family homes since 1970: new indexes for four cities. New England Economic Review 1987;(Sept/Oct); 45–56.

Case KE, Shiller RJ. The efficiency of the market for single-family homes. The American Economic Review 1989;79; 125–137.

Case B, Pollakowski HO, Wachter SM. On choosing among house price index methodologies. AREUEA Journal 1991; 19; 286–307.

Clapp MJ. A semiparametric method for valuing residential locations: application to automated valuation. Journal of Real Estate Finance and Economics 2003;27; 303–320.

Clapp JM, Giaccotto C, Tirtiroglu D. Housing price indices based on all transactions compared to repeat subsamples. AREUEA Journal, 1991;19; 270–285.

Clapp JM, Giaccotto C. Estimating price trends for residential property: a comparison of repeat sales and assessed value methods. Journal of Real Estate Finance and Economics 1992;5; 357–374.

Clapp JM, Giaccotto C. Price indices based on the hedonic repeat-sales method: application to the housing market. Journal of Real Estate Finance and Economics 1998;16; 5–26.

Clapp JM, Giaccotto C. Revisions in repeat-sales price indexes: here today, gone tomorrow? Real Estate Economics 1999;27; 79–104.

Goodman AC, Thibodeau TG. Dwelling age heteroskedasticity in repeat sales house price equations. Real Estate Economics 1998;26; 151–171.

Ono H., H.Takatsuji and C.Shimizu.Conjunct method of deriving a hedonic price index in a secondhand housing market with structural change, *Reitaku Institute for Political Economics and Social Studies Working Paper*, 2004;18.

Shimizu C, Nishimura KG. Biases in appraisal land price information: the case of Japan. Journal of Property Investment and Finance 2006;26; 150–175.

Shimizu C, Nishimura KG. Pricing structure in Tokyo metropolitan land markets and its structural changes: pre-bubble, bubble, and post-bubble periods," Journal of Real Estate Finance and Economics 2007; 35: pp.476-495.

Shimizu C, Nishimura KG, Asami Y. Search and vacancy costs in the Tokyo housing market. Regional and Urban Development Studies 2004; 16; 210–230.

Thibodeau TG. House Price Indices. , Kluwer Academic Publishers, 1997.