RIPESS

Reitaku Institute of Political Economics and Social Studies

Working Paper No.18

Conjunct method of deriving a hedonic price index in a secondhand housing market with structural change

Hiroya Ono[1], Hideoki Takatsuji[1] & Chihiro Shimizu[2]

Reitaku University, International School of Economics and Business Administration[1]

RECRUIT CO. LTD., Real Estate Research Institute[2]

November 25, 2004

RIPESS *Reitaku Institute of Political Economics and Social Studies*
Reitaku University
Hikarigaoka 2-1-1,Kashiwa, Chiba 277-8686
TEL+4-7173-3761/Fax+4-7173-3767

Conjunct method of deriving a hedonic price index in a secondhand housing market with structural change

Summary

Summary

A hedonic price index is generally defined as the ratio of the price of goods in the t^{th} period to the price in the base period, where these prices are estimated with conditions of structural change by using each period's hedonic price model. This paper, however, demonstrates that in the case of the Tokyo metropolitan secondhand housing market, the coefficients of each period's hedonic price model have moved up and down excessively over the whole estimation period for some reason other than structural changes. Therefore, the model would no longer be effective for the estimation of a hedonic price index. Instead, this paper attempts to develop an Overlapping Period Hedonic Model to remedy the defects of the hedonic price model as described above. Based on the assumption that structural change would take place gradually during a specified period, the new model is estimated by using the observations not only in a current period but also for recent periods of a specified length to eliminate the unexpected effects caused by reasons other than structural changes. A new price index, which is called here the *OPHM* index, is calculated from the new model.

1 Aim of this study

1.1 A housing price index and the change in housing quality issue

A hedonic price index is constructed to capture the price changes of goods and services between different points in time.[1] A hedonic price index is generally defined as the ratio of the price of goods or services in the t^{th} period and the corresponding price in the base period. In principle, it is necessary to compare goods and services of the same quality between different points in time. However, it is difficult to implement this principle in cases where the quality of goods and services in the market changes over time. Hence we need to adjust for changes to keep quality equivalent to that in the base period and then calculate the ratio of the two prices. This is particularly the case for those goods and services that have a variety of characteristics, and where judgment of quality relies on the consumer's preferences for those characteristics. A housing price index, our subject in this paper, is one of the typical examples of this practice.

It should firstly be noted that in constructing a housing price index, it is impossible to monitor the prices of houses of the same quality in the market repeatedly. This is because the specifications and facilities of each property differ. Even if they had very similar specifications, there would still be differences in the age of the buildings and their degree of depreciation. Residential property has strong individuality in its characteristics: in other words, it is heterogeneous. We would regard this as its "particularity with few equivalents". Secondly, the quality of residential properties, especially condominiums, changes in accordance with technical development through time. For example, walls and slab floors have been made thicker to improve

---

1 Typical examples are the Consumer Price Index and the Wholesale Price Index.

their sound insulation for greater protection of privacy. Other facilities such as floor heating, integrated kitchen systems and security systems have also improved rapidly.

1.2 The hedonic approach and its application to a housing price index

We have two approaches for coping with the above issue in constructing a housing price index. One is the Repeat Sales approach and the other is the Hedonic Pricing approach. The Repeat Sales approach focuses on sales transactions involving the same properties, which have happened repeatedly in the market. We can avoid the problem of particularity of the goods by observing the same sample of properties. However, there is no comprehensive database where we could observe repeated sales transactions of the same properties. Hence it is almost impossible for us to apply this approach to establish a housing price index.

In the hedonic approach, we assume that housing price (p) can be explained by the characteristics of each property. We would estimate a regression model representing housing price by using explanatory variables (z) such as commuting time to CBD, surrounding environment, floor space, condition of facilities, and age of the accommodation (These are preference criteria from the consumers' point of view).

In theory, consumers look for residential property with their preference criteria while housing suppliers try to provide housing that reflects those preferences. Given optimal behavior of both parties, we should be able to estimate a housing market price function ($p = p(z)$) with market equilibrium (Shepard 1999). This would be a hedonic price function. In the hedonic approach, we focus on housing that can be described by particular variables (z)—that is, housing of the same quality—and then compare its hedonic price between different points in time.

An index constructed by the hedonic approach is called a quality-adjusted price index or hedonic price index. In this paper, we estimate a hedonic price index for the secondhand housing market in the Tokyo metropolitan area.

1.3 The purpose of the study: a hedonic housing price index ($p = p(z)$)

Our starting issue is to build up a secondhand condominium price index for which we have n_t samples of sales transactions (p_t, z_t) for a period (t). To apply the hedonic approach, we need to estimate a hedonic price function ($p = p(z)$). We have two hypotheses for this application. One is to assume that the hedonic price function does not change during all periods of estimation. In this case, the coefficient of variance is the same for all periods. The estimated price function under this assumption is called a structurally restricted price model. The index established under this model is therefore a structurally restricted price index. Under the second hypothesis, the function allows for structural change during each period. This model is called a structurally nonrestricted price model. Accordingly, the index based on this function is a structurally nonrestricted price index (Nakamura, 1996).

In this paper, we start by building up these two types of hedonic price index and then investigate the following issues.

1. The characteristics of the two types of price index mentioned above and their relationship.
2. Positive analyses of the two indices.
3. The method of identifying structural change.
4. Problems and limitations of the two price indices with regard to structural change.

Finally, a new model for a hedonic housing price index (Overlapping Period Hedonic Model index) is proposed and its relevance examined.

2 Theoretical review of structurally restricted and nonrestricted price indices

Firstly, we review several characteristics of structurally restricted and nonrestricted price indices and the relationship between these indices. This is useful when we consider the issue of structural change in the later part of this section.

2.1 *Definition of a housing price index*

A housing price index is generally defined as the ratio of the price in the t^{th} period to the price in the base period.

$$Index_t = \frac{p_t}{p_0} \tag{1}$$

In our model, we make a logarithmic transformation of housing price p_t and variable $z_{t,i}$ (i^{th} variable as of period t) into $y_t = \log p_t$ and $x_{t,i} = \log z_{t,i}$ for analysis. Hence we amend equation (1) and define our housing price index, $Lindex_t$, in the following equation (2) for explanatory purposes.

$$\log Index_t = \log p_t - \log p_0 = y_t - y_0 \quad (= Lindex_t) \tag{2}$$

2.2 *Methodology: a structurally restricted price index*

Here we demonstrate how we construct a housing price index for four periods—period 0, period 1, period 2, and period 3—where the base point is period 0, housing price is y, and the only variable for explanation is x. For each period, the number of samples is n_t ($t = 0,1,2,3$) and each sample has a set of data on price and the explanatory variable (y, x).

The structurally restricted price index is established with a model that assumes that the regression coefficient does not change for all estimation periods. We add, however, time dummy factors—d_1, d_2, d_3—as constant terms in order to deal with price change as of period 1–3. The time dummy factor, $d_t\,(t=1,2,3)$, equals 1 where the period is t and 0 in other cases. Under the assumptions mentioned above, the model is:

$$y = a_1 d_1 + a_2 d_2 + a_3 d_3 + b_0 + b_1 x + u\,, \tag{3}$$

where:

$a_t\,(t=1,2,3)$ is the regression coefficient for the time dummy factor;

b_0 is a constant term;

b_1 is the regression coefficient for the variable x; and

u is the random error.

While we have n_t samples for each period, we construct a price model by using all data for the periods, since we assume that the regression coefficient does not change. The housing price under variable x can be estimated as:

$$\hat{y} = \hat{a}_1 d_1 + \hat{a}_2 d_2 + \hat{a}_3 d_3 + \hat{b}_0 + \hat{b}_1 x\,. \tag{4}$$

The housing price for each period is, with time dummy factor, expressed as:

$$\hat{y}_0 = \hat{b}_0 + \hat{b}_1 x$$

$$\hat{y}_1 = \hat{a}_1 + \hat{b}_0 + \hat{b}_1 x$$

$$\hat{y}_2 = \hat{a}_2 + \hat{b}_0 + \hat{b}_1 x \tag{5}$$

$$\hat{y}_3 = \hat{a}_3 + \hat{b}_0 + \hat{b}_1 x.$$

Consequently, the housing price index can be described as:

$$Lindex_0 = \hat{y}_0 - \hat{y}_0 = 0$$

$$Lindex_1 = \hat{y}_1 - \hat{y}_0 = \hat{a}_1$$

$$Lindex_2 = \hat{y}_2 - \hat{y}_0 = \hat{a}_2 \tag{6}$$

$$Lindex_3 = \hat{y}_3 - \hat{y}_0 = \hat{a}_3,$$

where the period 0 is the base point.

This means that the structurally restricted price index is determined by the regression coefficient of the time dummy factor, irrespective of the quality of housing expressed as x.

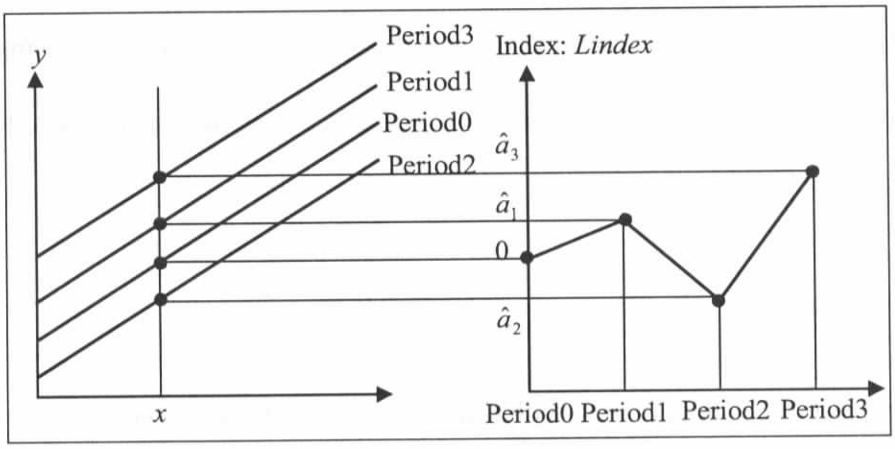

Figure 1. Structurally restricted price index.

2.3 *Methodology: a structurally nonrestricted price index*

With the structurally nonrestricted price index, we assume that the regression coefficient changes in each period. By using n_t samples for each period, the structurally nonrestricted price model is estimated as:

$$y_t = b_{t,0} + b_{t,1}x_t + u_t \ (t = 0,1,2,3), \tag{7}$$

where:

y_t is the housing price at period *t*;

$x_t \ (t = 1,2,3)$ is a variable expressing the quality of housing at period *t*;

$b_{t,0}$ is a constant term at period t;

$b_{t,1}$ is the regression coefficient at period *t*; and

u_t is the random error at period *t*.

The housing price for each period is expressed as:

$$\hat{y}_0 = \hat{b}_{0,0} + \hat{b}_{0,1}x$$

$$\hat{y}_1 = \hat{b}_{1,0} + \hat{b}_{1,1}x$$

$$\hat{y}_2 = \hat{b}_{2,0} + \hat{b}_{2,1}x \tag{8}$$

$$\hat{y}_3 = \hat{b}_{3,0} + \hat{b}_{3,1}x.$$

Consequently, the housing price index can be described as:

$$Lindex_0 = \hat{y}_0 - \hat{y}_0 = 0$$

$$Lindex_1 = \hat{y}_1 - \hat{y}_0 = (\hat{b}_{1,0} - \hat{b}_{0,0}) + (\hat{b}_{1,1} - \hat{b}_{0,1})x = c_1$$

$$Lindex_2 = \hat{y}_2 - \hat{y}_0 = (\hat{b}_{2,0} - \hat{b}_{0,0}) + (\hat{b}_{2,1} - \hat{b}_{0,1})x = c_2 \tag{9}$$

$$Lindex_3 = \hat{y}_3 - \hat{y}_0 = (\hat{b}_{3,0} - \hat{b}_{0,0}) + (\hat{b}_{3,1} - \hat{b}_{0,1})x = c_3$$

where the period 0 is the base point.

This means that the structurally nonrestricted price index depends on the variable x upon which we focus. Therefore, it is possible for a price index of a particular type of housing to go up while the index of another type of housing goes down.

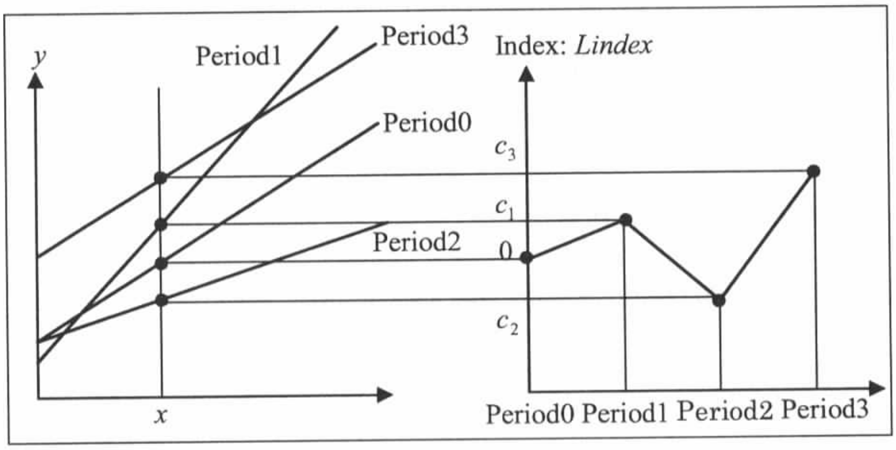

Figure 2. Structurally nonrestricted price index.

2.4 *Characteristics of structurally restricted and structurally nonrestricted price indices*

We summarized the characteristics of the two types of indices above, although some of them require only very basic regression analysis.

We now use a different representation. Assume that we work out a price index on the basis of two periods, namely period 0 and period 1 with the base point as period 0. Each period has price information of $y_0(n_0 \times 1)$ and $y_1(n_1 \times 1)$ as well as an explanatory variable of $X_0(n_0 \times K)$ and

$X_1(n_1 \times K)$ respectively. The description of $(n \times m)$ is a matrix (row vector $\times$ column vector). The number of observations is expressed as n_0 and n_1 for each period. Finally, K is the number of explanatory variables including the constant term.

Having all data for both periods together, the model is described as:

$$\tilde{y} = \tilde{X}\beta + \tilde{\boldsymbol{u}} , \qquad (10)$$

where:

$$\tilde{y} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}, \tilde{X} = \begin{pmatrix} \boldsymbol{0} & X_0 \\ \boldsymbol{1} & X_1 \end{pmatrix}, \beta = \begin{pmatrix} a_1 \\ \boldsymbol{b} \end{pmatrix}, \tilde{\boldsymbol{u}} = \begin{pmatrix} \boldsymbol{u}_0 \\ \boldsymbol{u}_1 \end{pmatrix}$$

1. $\boldsymbol{0}$ is a column vector of which all elements are zero.
2. $\boldsymbol{1}$ is a column vector of which all elements are one.
3. The first column of both X_0 and X_1 is one vector for the constant term.
4. The first column of $\tilde{X}$, $(\mathbf{0},\mathbf{1})'$ is a time dummy factor.
5. a_1 is the regression coefficient for the time dummy factor of period 1.
6. $\boldsymbol{b}(K \times 1)$ is a regression coefficient row vector for explanatory variables and the constant term, except for the time dummy factor.
7. $\beta(\overline{K+1} \times 1)$ is a regression coefficient vector that consists of the scalar of a_1 and the column vector of $\boldsymbol{b}(K \times 1)$.
8. Both $\boldsymbol{u}_0(n_0 \times 1)$ and $\boldsymbol{u}_1(n_1 \times 1)$ are random errors.

On the other hand, the model with structural change is described, by using the data for each period, as:

$$y_0 = X_0\beta_0 + u_0 \tag{11}$$

$$y_1 = X_1\beta_1 + u_1 . \tag{12}$$

Characteristic 1: The sum of estimation error for each period is zero.

In the case of the model with no structural change, $\tilde{y}$ is estimated as $\hat{y} = \tilde{X}\hat{\beta} = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$, where $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$. The error is described as $\hat{u} = \tilde{y} - \hat{y} = (I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')\tilde{y}$ accordingly. By multiplying $\tilde{X}'$ on the equation, we derive $\tilde{X}'\hat{u} = 0$. The first row vector of the data matrix is the time dummy factor for period 1 and the second row vector is 1 since it corresponds with the constant term. The equation of $\tilde{X}'\hat{u} = 0$, therefore, indicates that the sum of the estimation error for period 0 is zero and that the sum of the estimation error for period 1 is also zero.

Characteristic 2: The estimated equation $\hat{y} = \tilde{X}\hat{\beta}$ of the model with no structural change runs through the gravity point of observations of each period.

From the equation of $\hat{y} = \tilde{X}\hat{\beta}$, we have the following equations:

$$\hat{y}_0 = X_0\hat{b},\ \hat{u}_0 = \tilde{y}_0 - \hat{y}_0, \text{ therefore } \tilde{y}_0 - \hat{u}_0 = X_0\hat{b}, \tag{13}$$

and

$$\hat{y}_1 = \hat{a}_1 + X_1\hat{b},\ \hat{u}_1 = \tilde{y}_1 - \hat{y}_1, \text{ therefore } \tilde{y}_1 - \hat{u}_1 = \hat{a}_1 + X_1\hat{b} . \tag{14}$$

By multiplying $\boldsymbol{1}'$ (all vectors are 1) on both sides from the left and then dividing them by the observation numbers of each period, we have the following equations:

$$\bar{y}_0 = \bar{x}_0'\hat{b}, \tag{15}$$

and

$$\bar{y}_1 = \hat{a}_1 + \bar{\boldsymbol{x}}_1{}'\hat{\boldsymbol{b}}. \tag{16}$$

We find $\bar{y}_0$ and $\bar{y}_1$ to be the mean price of each period with $\bar{\boldsymbol{x}}_0$ and $\bar{\boldsymbol{x}}_1$ as averages of explanatory variance. Similarly, in the model with structural change, we have the following equations.

$$\bar{y}_0 = \bar{\boldsymbol{x}}_0{}'\hat{\boldsymbol{\beta}}_0 \tag{17}$$

$$\bar{y}_1 = \bar{\boldsymbol{x}}_1{}'\hat{\boldsymbol{\beta}}_1 \tag{18}$$

Characteristic 3: If the average of the explanatory variances of period 0 and period 1 are equal, i.e. $\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}_1$, and if the houses are of average quality, housing price indices derived from the two models are equal.

Where $\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}_1 = \bar{\boldsymbol{x}}$, the estimated housing prices with the value of $\bar{\boldsymbol{x}}$ are:

$$\bar{y}_0 = \bar{\boldsymbol{x}}'\hat{\boldsymbol{b}} = \bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}}_0 \tag{19}$$

and

$$\bar{y}_1 = \hat{a}_1 + \bar{\boldsymbol{x}}'\hat{\boldsymbol{b}} = \bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}}_1. \tag{20}$$

This means that the estimated housing prices with the value $\bar{\boldsymbol{x}}$ from both models are the average prices of each period.

2.5 *The relationship between the index with no structural change and the index with structural change*

In addition to equation (10) as well as (11) and (12), we add another model, which is the model with no structural change and without time dummy factors:

$$\tilde{y} = X\boldsymbol{\beta}_{*} + \boldsymbol{u}_{*}, \tag{21}$$

where

$$\tilde{y} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}, X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}, \boldsymbol{u}_{*} = \begin{pmatrix} \boldsymbol{u}_{*0} \\ \boldsymbol{u}_{*1} \end{pmatrix}.$$

$\boldsymbol{\beta}_{*}$ is the regression coefficient vector that includes the constant term but excludes time dummy factors.

Characteristic 4: If the variance of the random error of the two models is the same in all periods, the regression coefficient of the model with no structural change, $\hat{\beta}_{*}$, is the weighted average of the regression coefficients of the model with structural change, $\hat{\beta}_0$, $\hat{\beta}_1$, by the inverse matrix of the covariance matrix of the regression coefficient.

The estimated value of $\boldsymbol{\beta}_{*}$ in the equation (21) is:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{*} &= (X'X)^{-1}X'\tilde{y} \\ &= (X_0'X_0 + X_1'X_1)^{-1}(X_0'y_0 + X_1'y_1) \end{aligned} \tag{22}$$

The estimated values of β_0, β_1 in equations (11) and (12) are:

$$\hat{\beta}_0 = (X_0'X_0)^{-1}X_0'y_0, \tag{23}$$

and

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y_1. \tag{24}$$

Assuming that $u_0 \sim N(0, \sigma^2 I_0)$ and $u_1 \sim N(0, \sigma^2 I_1)$, namely that covariances of the random error terms are the same in each period, we obtain:

$$Var(\hat{\beta}_0) = \sigma^2 (X_0'X_0)^{-1} \; (= V_0), \tag{25}$$

and

$$Var(\hat{\beta}_1) = \sigma^2 (X_1'X_1)^{-1} \; (= V_1). \tag{26}$$

Equation (22) can then be altered by (23), (24), (25) and (26) so that:

$$\begin{aligned} \hat{\beta}_* &= (\sigma^2 V_0^{-1} + \sigma^2 V_1^{-1})^{-1} (\sigma^2 V_0^{-1} \hat{\beta}_0 + \sigma^2 V_1^{-1} \hat{\beta}_1) \\ &= (V_0^{-1} + V_1^{-1})^{-1} (V_0^{-1} \hat{\beta}_0 + V_1^{-1} \hat{\beta}_1) \end{aligned} \tag{27}$$

Equation (27) suggests that the regression coefficient of the model with no structural change, $\hat{\beta}_*$, is the weighted average of the regression coefficients of the model with structural change, $\hat{\beta}_0$, $\hat{\beta}_1$, by the inverse of the covariance matrix.

Similarly, as in the case of equation (10), which represents the model with no structural change including time dummy factors, the regression coefficients except for the constant and time dummy factors are the weighted averages of the regression coefficients of the model with structural change, $\hat{\beta}_0$, $\hat{\beta}_1$, by the inverse of the covariance matrix. This is simply because equation (10) is made by shifting the regression plane of equation (21) in parallel as it goes through the gravity point of observations for each period. Consequently, the regression coefficients of equations (10) and (21) are the same except for the constant term and time dummy factors.

Characteristic 5: The index with no structural change is the same as the index with structural change for those residential houses whose characteristics are defined as the weighted average of the average of explanatory variables for each period, weighted by the covariance matrix of the model with structural change in regression coefficients $\hat{\beta}_0$, $\hat{\beta}_1$.

As explained before, the index with no structural change for period 1 is represented as a time dummy factor $\hat{a}_1$ in equation (10). Thus, for period 1, the index with structural change for residential property with quality $\boldsymbol{m}$ can be described as:

$$\hat{y}_1 - \hat{y}_0 = \boldsymbol{m}'\hat{\beta}_1 - \boldsymbol{m}'\hat{\beta}_0 = \boldsymbol{m}'(\hat{\beta}_1 - \hat{\beta}_0)\,. \tag{28}$$

We can solve this equation to find out the value of $\boldsymbol{m}$ as the index equals $\hat{a}_1$. This helps us understand under which conditions the two indices show the same movement.

Provided that $\bar{y}_0$, $\bar{\boldsymbol{x}}_0$, $\bar{y}_1$ and $\bar{\boldsymbol{x}}_1$ are average prices and values of each period, the index with no structural change ($\hat{a}_1$) is:

$$\begin{aligned}\hat{a}_1 &= (\bar{y}_1 - \bar{\boldsymbol{x}}_1'\hat{\beta}_*) - (\bar{y}_0 - \bar{\boldsymbol{x}}_0'\hat{\beta}_*) \\ &= (\bar{y}_1 - \bar{y}_0) - (\bar{\boldsymbol{x}}_1' - \bar{\boldsymbol{x}}_0')\hat{\beta}_*\end{aligned} \tag{29}$$

This equation can be transformed by $\hat{\beta}_* = (V_0^{-1} + V_1^{-1})^{-1}(V_0^{-1}\hat{\beta}_0 + V_1^{-1}\hat{\beta}_1)$, which is derived from equation (27) into:

$$\hat{a}_1 = (\bar{\boldsymbol{x}}_0'W_0 + \bar{\boldsymbol{x}}_1'W_1)(\hat{\beta}_1 - \hat{\beta}_0)\,. \tag{30}$$

We know that:

$$W_0 = V_0(V_0 + V_1)^{-1}, \tag{31}$$

and

$$W_1 = V_1(V_0 + V_1)^{-1}. \tag{32}$$

Therefore, we can work out the value of $\boldsymbol{m}$ where $\boldsymbol{m}'(\hat{\beta}_1 - \hat{\beta}_0) = \hat{a}_1$ as follows:

$$\boldsymbol{m}' = \bar{x}_0' W_0 + \bar{x}_1' W_1. \tag{33}$$

This shows that $\boldsymbol{m}$ is the weighted average of the average of the explanatory variables for each period, weighted by the covariance matrix of the regression coefficients $\hat{\beta}_0$, $\hat{\beta}_1$ of the model with structural change. This also demonstrates the case for Characteristic 3 because $\boldsymbol{m} = \bar{x}$ if $\bar{x}_0 = \bar{x}_1 = \bar{x}$.

3 Positive analysis of the price index with no structural change and the price index with structural change

3.1 Basic model

The basic equation of the model with no structural change is described as follows: the explanatory variables for residential property characteristics (z) include floor space (*FS*), distance from the nearest station (*WK*), accessibility to the CBD (*ACC*), age of building (*BY*), area of balcony (*BS*) and other building characteristics (BC_h). As for location factors, there is a railway dummy factor (RD_i) and an administrative area dummy factor (LD_j). Finally, we have a time dummy factor (TD_k) as well. The regression coefficient of the time dummy factor (a_{12k})

represents the secondhand condominium price index. The application of this model to pooled data from period 1 to period T makes an estimate of the index with no structural change.

$$\begin{aligned}\log RP_g &= a_0 + a_1 \log WK + a_2 \log ACC + a_3 \log FS \\ &+ a_4 \log BY + a_5 \log BS + a_6 \log NU + a_7 \log NR + a_8 RT \\ &+ \sum_h a_{9,h} \cdot BC_h + \sum_i a_{10,i} \cdot RD_i + \sum_j a_{11,j} \cdot LD_j + \sum_k a_{12,k} \cdot TD_k + \varepsilon\end{aligned} \tag{34}$$

RP	:	*Price of Condominium*
WK	:	*Distance to nearest station*
ACC	:	*Accessibility to Central Business District*
FS	:	*Floor Space/Square Meters*
BY	:	*Number of Years After Construction*
BS	:	*Balcony Space/Square*
NU	:	*The Number of Units*
RT	:	*Market Reservation Time*
BC	:	*Other Building Characteristics*
RD	:	*Rail Dummy*
LD	:	*Local Dummy*
TD	:	*Time Dummy*

N.B. We do not investigate the case for $g = 2$: condominium rent in this paper.

The equation of the model with structural change can be described as the model with no structural change by removing the time dummy factor, although the model must be applied to data for each observed period. We then have 159 models from 159 periods.

$$\begin{aligned}\log RP_{gt} &= a_0 + a_{1,t} \log WK_t + a_{2,t} \log ACC_t + a_{3,t} \log FS_t \\ &+ a_{4,t} \log BY_t + a_{5,t} \log BS_t + a_{6,t} \log NU_t + a_{7,t} \log NR_t + a_{8,t} RT_t \\ &+ \sum_h a_{9,h,t} \cdot BC_{h,t} + \sum_i a_{10,i,t} \cdot RD_{i,t} + \sum_j a_{11,j,t} \cdot LD_{j,t} + \varepsilon_t \qquad (t = 1 \cdot \cdots \cdot \mathrm{T})\end{aligned} \tag{35}$$

3.2 *Data for analysis*

3.2.1 *Secondhand condominium price data*

The database in Ono et al. (2001) has been updated for this analysis. The main information resource is *Residential Information Weekly* (or *Shukan Jhutaku Joho* in Japanese). From this, we can gain information on characteristics and asking prices of listed property on a weekly basis. Hence there are historical price data of individual properties from when they are put on the market in the magazine for the first time until they are removed owing to sale or other reasons. The most important information is the original asking price that appears in the magazine/market, the final asking price in the magazine/market before it is removed and the actual sales price (although actual sales prices are not all available in the magazine).

The original asking price represents the seller's desired price rather than the market value. Some of the actual sales prices reflect specific situations of individual transactions such as "quick sale" or "hasty purchase", which are not uncommon but quite typical in property transactions. Therefore, we have collected the final asking prices from the magazine and analyzed the samples. The last asking price shown in the magazine is the first bid price offered by a prospective buyer. This happens through the process where several particulars of a residential property are disclosed to the market in the magazine and the buyer responds to that information. Thus the price indicates an upper range of possible bid prices and can be regarded as a competitive market price, relatively free from individual specific conditions associated with transactions.

3.2.2 Data on characteristics of residential property

With regard to building structure, we excluded samples of condominiums with steel frame structure from our database. This is because almost all residential investment properties are either of reinforced concrete or steel-reinforced concrete structure.

The transport accessibility of each location is expressed by "distance from the nearest station" (*WK*) and "accessibility to CBD" (*ACC*). *ACC* is measured in the following way. First, we designated the 40 biggest train terminal stations in terms of the number of passengers in the metropolitan area as the "CBD" since the model is designed for this area. Secondly, we investigated time distances from all stations in the area to the 40 terminal stations and weighted them by the number of passengers for each terminal station to calculate the ACC for each station.[2] The *ACC* shows the average time distance of each station to the terminals.[3] By considering the time distances to multiple terminal stations, the time-saving effect from new railway developments on the whole transport network can be embedded. If the property is located within walking distance of the CBD, we use a dummy factor, *WD*, to show it.

Transaction price is also influenced by disparity between demand and supply. This can be represented by marketing time (RT), which starts when the seller first puts a property on to the market and ends when the property sells. A long *RT* suggests that there was a big gap between the seller and the buyer in terms of price. A short *RT* implies that the first asking price was nearly market price and that it was easy for the seller and buyer to agree on the price. We measured how long it took for a property placed in the magazine to be removed owing due to transaction and recorded it in a database.

We also identified a few quantitative measures representing building characteristics. They are floor space (*FS*), age of building (*BY*), balcony space (*BS*), and the number of units (*NU*). The number of units is regarded as a proxy variable for the grade of the property and the quality of the common space.

We set other types of dummy factor. The *ground floor dummy factor* is used since the price of a ground floor property tends to be lower than that of higher floor properties, except for those that have exclusive gardens. A *top floor dummy factor* is employed as well. In terms of direction of property, there are two dummy factors, namely the *south-facing dummy factor* (*SD*) and the

---

2 We use a database provided by Jordan Ltd.

3 In order to undertake monthly analysis, it is desirable to renew ACC as soon as railway timetables change. However, we made changes on 1 April and 1 October each year.

south-directed dummy factor, which includes southeast and southwest facing properties. Furthermore, we have the *RSC dummy factor* and the *HLC dummy factor* as proxies for housing quality. The RSC dummy factor indicates that the property is of reinforced concrete steel structure while the HLC dummy factor suggests that the Housing Loan Corporation mortgage is available for the property.

The above variables are all related to location and building characteristics. It is reasonable to assume that wider regional factors could also affect the housing price. Therefore, we set an *administrative area dummy factor* (*LD*) to reflect differences in the quality in public services, as well as *Ji-Gurai* or area rank. Finally the *railway dummy factor* (RD) represents historical background in that most of the Japanese residential developments were implemented along railway lines.

Table 1 describes these data while Table 2 summarizes key statistics for each variable.

Table 1. List of analyzed data.

Symbols	Variables	Contents	Unit
WK	Distance to nearest station	Time distance to the nearest station (walking time and bus time).	minute
ACC	Accessibility to central business district	Average of railway riding time in daytime to the most crowded 40 stations in 1988 weighted by the number of passengers at the stations*.	minute
FS	Floor space/ square meters	Floor space (as shown in *Jutaku Joho* magazine**).	m^2
BY	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.	year
BS	Balcony space/ square meters	Balcony space (as shown in *Jutaku Joho* magazine).	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.	date
MC	Management cost	Management fee.	YEN/ month
WD	Walk dummy	Whether the time distance includes riding time of bus 1, not including bus time 1 including bus time 0.	(0,1)
FF	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)
SD2	South-facing dummy2	Fenestrae facing south, south west or south east 1, other directions 0.	(0,1)
TK	Ferroconcrete dummy	Steel reinforced concrete frame structure 1, other structure 0.	(0,1)
KD	Housing Loan Corporation dummy	Eligible for Housing Loan Corporation loan 1, not eligible 0.	(0,1)
RDi (i=0,...,I)	Railway line dummy	*i th railway line 1, other railway line 0.* (10 railway lines appeared in the magazine)	(0,1)
LDj (j=0,...,J)	Location (Ward) dummy	*j th administrative district 1,* other district 0.	(0,1)
TDk (k=0,...,K)	Time dummy (monthly)	*k th month 1,* other month 0.	(0,1)

* Shinjhuku station is the busiest station. The busiest 40 stations include main terminal stations of Yamanote Line such as Shinagawa, Ikebukuro and Shibuya as well as Yokohama, Kawasaki, Chiba, Omiya and Kashiwa stations. We have established a 73,920 railway line network database, which is worked out of 1848 stations appeared in the magazine for the 40 stations. This database is updated every six months.

** a weekly residential listing magazine by RECRUIT

Table 2. Key statistics of secondhand condominium price data.

Variables	Average	Standard Deviation	Minimum	Maximum
RP: Secondhand condominimum price (10,000 Yen)	4,199.54	3,018.12	500.00	36,300.00
FS: Floor space (m²)	55.30	18.85	15.00	120.00
RP/FS	75.59	44.02	12.14	695.08
WK: Distance to the nearest station (minutes)	7.68	4.32	1.00	26.00
ACC: Accessibility to Central Business District (minutes)	25.14	5.02	16.31	92.69
BY: Age of building (year)	13.74	7.08	2.00	35.00
BS: Balcony Space (m²)	7.33	5.85	5.00	165.20
NU: The Number of Units	101.87	132.08	6.00	1324.00
RT: Market reservation time (day)	88.96	86.55	1.00	700.00

1989/04-2002/06 n=164,088

3.2.3 Observation period and monthly data

The number of samples is 164,088 in the Tokyo Special District (23 wards) between April 1989 and June 2002. The model is estimated on a monthly basis. There are 159 observation periods.

3.3 The results of the index with no structural change estimate

The results of our estimate of a residential price index with no structural change (*RHI: Resale House Index*) are shown in detail in Table 3. The degree of precision of the model is good, as shown by the coefficient of determination adjusted for the degrees of freedom.

In Figure 3, the *RHI* is compared with the average value index (*AI:* the ratio of the average price calculated monthly to the average price in April, 1989), and the monthly number of samples was shown simultaneously. By *RHI* and *AI*, it turns out that a different trend occurred. Furthermore, a large variation is evident in the monthly number of samples.

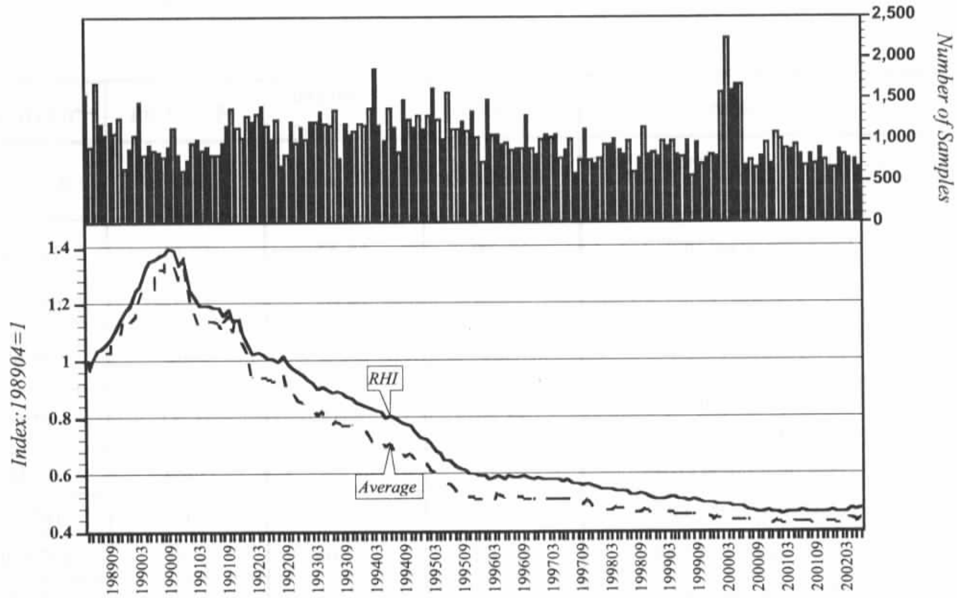

Figure 3. Secondhand residential price index with no structural change: *RHI*, *AI* comparison: 1989/04 - 2002/06.

Table 3. Estimated results of the model with no structural change: Tokyo's 23 wards.

Method of Estimation

OLS

Dependent Variable

RP: Resale Price of Condominiums (in log)

Independent Variables

Property Characteristics (in log)	Coefficient	t-value
Constant	6.182	327.650
FS: Floor space	0.998	639.660
BS: Balcony space	0.021	24.255
NU: Number of units	0.023	41.768
RT: Market reservation time	0.021	41.982
ACC: Accessibility to CBD	-0.424	-77.887
WK: Distance to the nearest station	-0.054	-70.975
BY: Age of building	-0.186	-266.206
Property Characteristics (dummy variables)		
WD: Walking distance	0.122	29.920
SD: South-facing	0.008	7.312
KD: Housing Loan Corporation	0.068	19.872

Ward (city) Dummy RDi $(i=0,\ldots,I)$	Coefficient	t-value
Chiyoda	0.331	55.457
Minato	0.159	50.440
Shinjhuku	0.046	15.406
Taito	-0.303	-64.684
Sumida	-0.302	-72.409
Koto	-0.316	-99.554
Shinagawa	-0.046	-13.318
Meguro	0.035	8.879
Ota	-0.047	-12.893
Setagaya	0.085	27.799
Shibuya	0.198	59.443
Nakano	-0.036	-9.962
Toshima	-0.087	-23.965
Kita	-0.190	-33.093
Arakawa	-0.395	-91.019
Itabashi	-0.192	-46.598
Nerima	-0.084	-22.497
Adachi	-0.432	-95.023
Katsushika	-0.366	-71.098
Edogawa	-0.280	-72.991

Railway/Subway Line Dummy LDj $(j=0,\ldots,J)$	Coefficient	t-value
Yamanote Line	0.042	17.400
Ginza Line	0.166	32.650
Marunouchi Line	0.017	5.041
Hibiya Line	0.126	36.642
Chiyoda Line	0.042	11.271
Yurakucho Line	-0.010	-3.016
Hanzomon Line	0.046	5.342
Nanboku Line	-0.037	-3.649
Toei Mita Line	-0.042	-11.044
Toei Shinjuku Line	-0.012	-4.331
Yurikamome	-0.318	-2.326
Tokyo Monorail	-0.147	-16.655
Keihin Kyuko Line	-0.175	-45.615
Kuko Line	-0.161	-18.559
Keihin Tohoku Line	-0.024	-5.902
Tokyu Ikegami Line	0.042	8.950
Tokyu Oimachi Line	0.034	7.618
Tokyu Toyoko Line	0.048	12.590
Tokyu Shin-Tamagawa Line	0.069	5.072
Tokyu Setagaya Line	-0.079	-10.296
Odakyu Line	-0.009	-2.440
Keio Inokashira Line	0.024	4.795
Keio Shi-Line	-0.128	-39.617
Chuo Line	0.026	7.545
Seibu Shinjhuku Line	-0.029	-7.345
Seibu Ikebukuro Line	-0.058	-14.159
Tobu Tojyo Line	-0.047	-9.976
Saikyo Line	-0.128	-16.912
Takasaki Line	-0.074	-11.465
Tobu Isezaki Line	-0.015	-3.143
Joban Line Express	0.021	3.641
Keisei Oshiage Line	-0.058	-9.710

Adjusted R square= 0.884

Number of Observations= 164,088

3.4 The results of the price index with structural change estimate

The estimated results of the model with structural change can be explained as follows. As explanatory factors, we used several residential characteristics that are shared during all the periods to estimate secondhand condominium prices for each period. The price index with structural change was constructed from the base period onwards. Figure 4 shows the price index for Tokyo's 23 wards. The broken line labeled '*NRHI*' indicates the price index with structural change while the solid line named '*RHI*' is the line for the price index with no structural change.

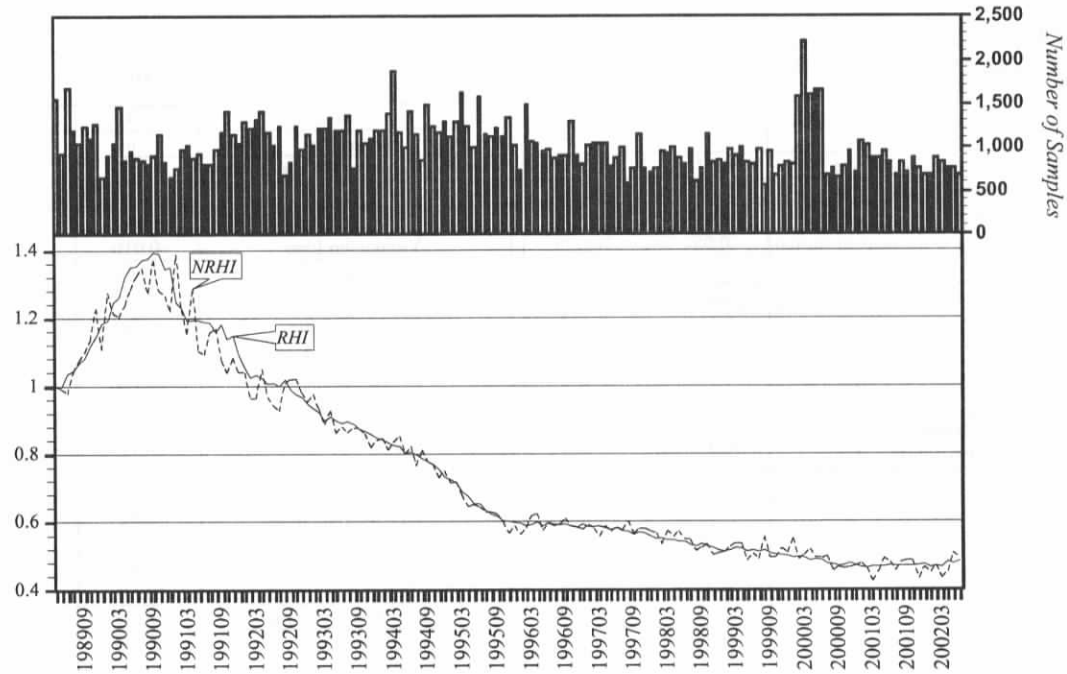

Figure 4. Comparison of price indices with structural change (*NRHI*) and no structural change (*RHI*): 1989/04 - 2002/06.

3.5 The degree of volatility of the index with structural change

Comparing the two indices shown in Figure 4, we can see that the price index with structural change is more volatile than the index with no structural change. This fact does not seem to match our perception of real housing price movements. For instance, the graph suggests that housing prices fell significantly in a particular period. However, this phenomenon has not been observed in the market where residential properties with particular characteristics have suffered from large price falls in any period. Also, it is hardly possible to imagine that residential housing prices have changed as drastically every month as the line suggests.

While we cannot judge *a priori* the superiority of an index by the degree of volatility of the index, the price index with structural change is certainly a method of indexation for prices with structural change. If we put this index into practical use, we would need to find out the case for such a drastic movement of the index. Otherwise, we have to assume that this change happens in the statistical inference and work out how we can have an index without such movement.

Therefore, we put the issue of index superiority aside and focus on the degree of volatility of the price index with structural change to investigate why it happens.

3.6 *Accuracy of estimates*

Figure 5 below shows the number of samples and adjusted coefficient of determination for the 159 estimates by the model with structural change for each period.

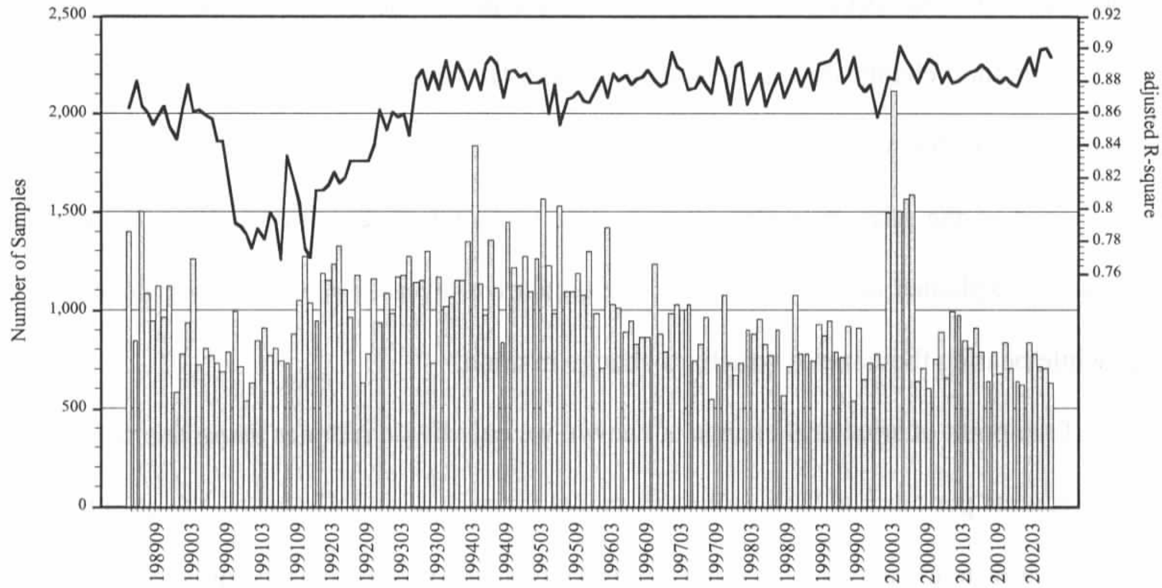

Figure 5. Accuracy of the model with structural change: Monthly: 1989/04 - 2002/06.

1. The adjusted coefficient of determination is 0.85–0.90 with the exception of the period 1991–1992 (approximately 0.80), which is very good.

2. The number of samples in several months is twice as many as in the other months. However, the sample size shows little relationship with the coefficient of determination. In other words, the number of samples has not affected the coefficient of determination.
3. We can identify, in Figure 3, some periods showing a large movement in the NRHM index between 1991 and 1992, when the coefficient of determination is not very good. However, other periods with good explanatory power also show great volatility.
4. Consequently, the explanatory power of the models has influenced the volatility of the index.

4 The Overlapping Period Hedonic Model index in the secondhand housing market

4.1 Issues

The price index with no structural change is a hedonic price index based on the assumption that there is no structural change during the subject period. However, this assumption is unreasonable if we consider our observation in the last section where it is highly likely that structural changes have happened in the past. It is also unreasonable to switch to the price index with structural changes as we explained before. Therefore, we summarize several issues regarding housing price indices while bearing the issue of structural change in mind.

1. If the point of structural change is known, we can divide samples using this point into separate periods and then estimate a model for each period. However, the question is not how to identify the breaking point(s) when we know neither the number of points nor when they are likely to be. This is the first question to be answered for structural change.

2. Even if we know the breaking point(s), they only explain structural change in the past. We would face the situation where we need to import information from a set of new data on monthly basis. The second question, therefore, is how to estimate a monthly structure that might change because its underlying database is updated every month.
3. Our purpose is not to estimate a price model with structural change but to establish a price index. The third issue is how to connect price indices with structural changes.

In the next section, we investigate several issues on structural change and develop our new methodology for price indices.

4.2 Issues of structural change

The preferences of homebuyers for location and housing, such as "distance to the nearest station" and "occupied space," will change over time. In addition, external factors such as the tax regime have impacts on behavioral patterns in housing selection. These affect the changes in regression coefficients in the price models. The methodology applied for detection of structural changes is to estimate regression models for the pre- and postbreaking points separately and then examine the equality of regression coefficients. This is called a structural change test.

In this case, there are the following problems for a structural change test. If we know that a particular external "shock" has created a break point, we can carry out a structural change test around that period. However, we cannot specify the month of a break point in advance. The break point can be the exact month in which an external shock happens or it can be two or three months later. We cannot tell how many break points we have over the whole period either. In technical terms, it is likely that error variance is uneven from month to month, while the variance itself is unknown.

Thus, we assume that the structure changes every month and estimate a price model for each in order to avoid the problem of unknown break points. If there is no structural change, the models should have equal regression coefficients. If there is a structural change, the regression coefficients will change in that month. At first, this approach seems to work, but it produces huge volatility in the regression coefficients for each month of our estimate (see 4.3). While this result suggests that there are structural changes, we cannot conclude that the structure changes every month from this observation. It can be assumed that the large volatility of the regression coefficients is caused by sampling bias. Therefore, it is unacceptable for us to use individual models.

Consequently, we believe that the best way is to estimate models for each period after dividing the period and data into suitable segments with break points. We already have plenty of existing studies regarding the methodology of structural change tests and so can undertake the test.[4] In particular, the best subset selection procedure is the most useful when we do not know when and how many break points we have. Assuming that there is only one break point, we would undertake a structural change test for each month in turn. Then we would do the same for all possible combinations of two months based on the assumption that there are two break points. We would also do the same for all the other possible combinations. This is far from efficient, but it is a manageable method.

---

4 The structural change test is an equality test of regression coefficients β_1, β_2. The methodology of the equality test depends on the assumption of the variances of the error term; i.e., either $\sigma_1^2 = \sigma_2^2$ or $\sigma_1^2 \neq \sigma_2^2$. When the variances of the error term are equal ($\sigma_1^2 = \sigma_2^2$), we have a general method for verifying the linear hypothesis on regression coefficients. If the variances of the error term are not equal ($\sigma_1^2 \neq \sigma_2^2$), we can use the asymptotic likelihood ratio test and calculate the unknown coefficient by using the fact that $-2log$ (skewness ratio) shows a chi-square distribution. Amemiya (1985) introduces in detail the method of dealing with this issue.

In conclusion, the issue that the structural change test arrives at is how to select a suitable model from a band of estimated regression models, estimated for alternative combinations of the subject periods. The more break points we have, the better the model fits the estimation, but we need more variables in total. Because of this trade-off situation, we focus on the balance between the goodness of fit of the model and the number of variables. We can then apply AIC to select the most suitable model.[5]

4.3 *Estimate of successively changing structure*

Suppose that our structural change tests reveal break points (t_1, t_2, ..., t_k) for the observed period (from 0^{th} to T^{th}). We then face the following questions. Will the structure of the period from the most recent break point t_{kt} to the current period of T remain the same for the next period of T+1? Alternatively, will we need to undertake another structural test between the pooled data covering the periods from t_k to T^{th} and the new set of data for $T+1^{th}$? If we do find a structural change, do we need to estimate a new model for the period T+1? Does the estimation of a new model for each new period in the following periods mean that the model becomes one with structural change? These questions suggest that our next issue is how to incorporate a new structure with additional data into the existing model constructed from the previous set of data.

This issue can be paraphrased in the following way. First, we think that structural change does not occur and finish at a break point instantaneously. Instead, it is likely that some time is needed for adjustment to a change caused by an external shock to be manifested. Thus the regression

5 With regard to this issue, Garcia and Perron (1996) state the methodology of the structural change test for the case where we know that structural change happens twice but do not know when the break points are. Jushan and Perron (1998) discussed the case where we do not know how often and when structure changes. Takatsuji (2001) proposed a method of exploratory analysis of structural change by using discrete time dummy factors and the AIC assuming that variance is even. This method is applied to the analysis of similar data in Nishimura and Shimizu (2002), assuming that there are two break points around a bubble period.

coefficient also changes successively and not at once. In this case, an identified break point tells us the time when successive changes become too great to ignore from a statistical point of view. However, this does not necessarily mean that we can predict a "true" regression coefficient that changes successively by estimating individual regression models for each of the break points.

The estimate of a model with structural change is one issue and the estimation of successively changing "true" regression coefficients is another. In the model with structural change, we disregard all data belonging to the period prior to the break point. Thus, we cannot evaluate the successive changes. Consequently, the issue can be restated as follows.

How can we incorporate a successively changing regression coefficient with additional data into the existing model (constructed from the previous sample) on the assumption that the structural change occurs successively?

We have two methods for dealing with this issue. One is the application of a Kalman filter and the other is the Overlapping Period Hedonic Model, which we will propose later.

4.4 Kalman filter

A discrete-time Kalman filter is a method for estimating successively changing regression coefficients where there are regularly updated observations.

We assume that the regression coefficient as of period t is expressed as an unknown state variable $\beta_t (K \times 1)$ [6], the observed secondhand condominium price data is $y_t (n_t \times 1)$ and the explanatory variable is $X_t (n_t \times K)$. More precisely, both the price data and the explanatory variable are transformed into logarithms. The number of observations as of period t is described as n_t while K is the number of explanatory variables including the constant term.

---

6 We express the matrix of m rows and n columns as (*man*). Any vectors x describe a row vector while x' represents a column vector.

Additionally, we set an assumption that, where unknown, the regression coefficient β_t is equal to the previous regression coefficient β_{t-1} and changes successively due to external shock (random error $v_t(K \times 1)$). Hence we can describe these conditions as follows.

$$\beta_t = \beta_{t-1} + v_t \tag{36}$$

$$y_t = X_t \beta_t + u_t \tag{37}$$

where $u_t(n_t \times 1)$ is the random error for the price data.

We can rewrite this issue by using these symbols as follows.

How can we incorporate a successively changing regression coefficient (β_t) with additional data (y_t) into the existing model (β_{t-1}) (constructed from the previous sample) on the assumption that the structural change occurs successively?

Suppose that we know a model β_{t-1} for the period $t-1$, which is a chance variable with an average and variance. β_t and y_t for the period t are functions of chance variables, β_{t-1}, v_t, u_t. The joint probability density is described as:

$$f(\beta_t, y_t) = f(\beta_t \mid y_t) f(y_t). \tag{38}$$

On the right-hand side, the first term, $f(\beta_t \mid y_t)$, means the conditional probability density of β_t where we have y_t. In other words, this term is a formulation of the issue mentioned above, namely "How can we incorporate a successively changing regression coefficient (β_t) with additional data (y_t) into the existing model (β_{t-1}) constructed from the previous sample?" Based on this conditional probability density, we can solve this formula by maximum likelihood estimation of β_t.

$$f(\beta_t \mid y_t) \to \max_{\beta, v, u} \tag{39}$$

From equation (38), $f(\beta_t \mid y_t)$ can be described as

$$f(\beta_t \mid y_t) = \frac{f(\beta_t, y_t)}{f(y_t)} = \frac{f(\beta_t, y_t)}{\int f(\beta_t, y_t) d\beta_t}. \tag{40}$$

Generally, the Kalman filter equation is sought from the variance–covariance matrix of v_t and u_t. We do not know the matrix, though, and hence we need to estimate it together with β_t by the maximum likelihood method. While there is a technical difficulty, the estimation would be easier by assuming that the variance–covariance matrix of v_t and u_t is a diagonal matrix.

This application of the Kalman filter is a potent method for estimation of successively changing regression coefficients[7]. However, we are not sure about characteristics of this solution for our particular data and hence put aside the application of the Kalman filter as our task for the future.

4.5 An Overlapping Period Hedonic Model

Generally, we estimate hedonic models with structural change by observing a set of data in certain periods separated by each break point. In other words, we break the connection of the data at the break points. This makes it difficult for us to identify regression coefficients with the assumption that structural change happens successively. Thus it would be reasonable to estimate the regression coefficients, in a similar way to that in which we would work out a moving average, by establishing consecutive hedonic models with a specified period.

---

7 Harvey, A.C. (1989), Maddala, G.S. and Kim, In-Moo (1998).

We backdate by τ periods and estimate $\boldsymbol{\beta}_t$ with those samples between period $t-\tau+1$ and period t. The model is:[8]

$$\tilde{\boldsymbol{y}}_t = \tilde{X}_t \boldsymbol{\beta}_t + \tilde{\boldsymbol{u}}_t , \tag{41}$$

where

$$\tilde{\boldsymbol{y}}_t' = (\boldsymbol{y}_t', \boldsymbol{y}_{t-1}', \cdots, \boldsymbol{y}_{t-\tau+1}')$$

$$\tilde{X}_t' = (X_t', X_{t-1}', \cdots, X_{t-\tau+1}')$$

$$\tilde{\boldsymbol{u}}_t' = (\boldsymbol{u}_t', \boldsymbol{u}_{t-1}', \cdots, \boldsymbol{u}_{t-\tau+1}') .$$

Then we work through the following process.

1. The initial period is $t = \tau - 1$. (We shall change this figure later.)
2. Estimate the model based on samples from period $t-\tau+1$ to period t.
3. Move forward to period $t+1 \rightarrow t$ and repeat step 2 until the current period.

In this estimation, samples are repeatedly used in a specified period for τ. This is the way by which we can estimate a successively changing structure. We do not break connectivity between the set of samples at break points but find out their structural change by connecting individual models consecutively.

We call this the Overlapping Period Hedonic Model (*OPHM*) and call period τ as the "overlapped estimate period". We estimate our model in this way. Our secondhand price index is based on the *OPHM* model as we explain in detail in the next section.

---

8 $\mathbf{x}'$ and A' represent transposition of vector $\mathbf{x}$ and matrix A respectively.

4.6 The Overlapping Period Hedonic Model index: a proposal

We apply *OPHM* to a hedonic index, *OPHI*. The *OPHM* is a model that has a particular structure for a specified period of τ. This means that we can estimate a price index in the form of time dummy coefficients for the time dummy factors for the observed period; i.e., $\tau - 1$ time dummy factors for τ periods. In *OPHM*, we estimate the model by moving the application period of τ month by month. The key point is how we can connect the indices.

We develop *OPHM* for period t based on formula (15) as shown below.

$$\tilde{y}_t = \tilde{X}_t \beta_t + \tilde{u}_t \tag{42}$$

where

$$\tilde{y}_t = \begin{pmatrix} y_{t-\tau+1} \\ \vdots \\ y_{t-1} \\ y_t \end{pmatrix}, \tilde{u}_t = \begin{pmatrix} u_{t-\tau+1} \\ \vdots \\ u_{t-1} \\ u_t \end{pmatrix}, \boldsymbol{\beta}_t = \begin{pmatrix} a_{t-\tau+2} \\ \vdots \\ a_{t-1} \\ a_t \\ \mathbf{b}_t \end{pmatrix} \quad however\,(\overline{\tau - 1 + K} \times 1),$$

$$\tilde{X}_t = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & X_{t-\tau+1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & & \mathbf{0} & X_{t-\tau+2} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & & \mathbf{0} & X_{t-\tau+3} \\ \vdots & & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} & X_{t-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} & X_t \end{pmatrix} \quad however\,(\overline{n_{t-\tau+1} + \cdots + n_{t-1} + n_t} \times \overline{\tau - 1 + K}).$$

This assumes that we estimate an *OPHM* for t periods and time dummy factors of $\hat{a}_{t-\tau+2}, \cdots, \hat{a}_{t-1}, \hat{a}_t$. These are price indices of which the base point is $t - \tau + 1$ while the indices cover $t - \tau + 2$, ..., $t - 1$ and t.

Then we focus on:

$$\hat{a}_t - \hat{a}_{t-1}. \tag{43}$$

This shows the difference in the price index between periods t and $t-1$. We then define this as the difference in the price indices for all periods, namely between $Lindex_t$ for period t and $Lindex_{t-1}$ for period $t-1$.

$$Lindex_t - Lindex_{t-1} = \hat{a}_t - \hat{a}_{t-1} \tag{44}$$

We can establish our price index for all periods as follows.

1. Try to establish the price index, $Lindex_{t-1}$, for the period up to $t-1$.
2. Have additional sample for period t.
3. Estimate an *OPHM* from the data between period $t-\tau+1$ and period t.
4. Have time dummy factor coefficients, $\hat{a}_{t-\tau+2}, \cdots, \hat{a}_{t-1}, \hat{a}_t$, corresponding to periods between $t-\tau+2$ and t.
5. Establish the price index, $Lindex_t$, for period t by application of $\hat{a}_t - \hat{a}_{t-1}$.

$$Lindex_t = Lindex_{t-1} + (\hat{a}_t - \hat{a}_{t-1}) \tag{45}$$

6. Repeat the same process for the next period.

5 Positive analysis of the *OPHM* index

5.1 How to decide the overlapped estimate period

It is very important to decide the overlapped estimate period of τ (see section 4.5) for making a Overlapping Period Hedonic Model index. We need to pay careful attention to the following points.

1. The purpose of setting the overlapped estimate period is to exclude effects on the regression coefficients caused by unique sampling bias in the monthly data if we separate our sample on a monthly basis. It is impossible for us to distinguish these effects from structural change. We do not know the reason for the bias (if any) at this stage. However, we would like to exclude seasonal changes of the market at least, and we know the season when many people move. Therefore, the overlapped period should be no less than one year since we need to cover all four seasons.
2. It would be unreasonable to think that the structural change spreads into the market in a single month. There should be an adjustment period for the structural change. Therefore, the overlapped estimate period should be long enough to cover this adjustment period.
3. Conversely, we would miss the structural change if we made the overlapped estimate period too long. We need to keep the period short enough so as not to miss the change.

Consequently, we have no clear criteria for setting the overlapped estimate period and hence use three years or 36 months for the period at the moment.

5.2 Characteristics of OPHM *index*

The results from estimation using the Overlapping Period Hedonic Model index are as follows.

1. The index shows similar movement to the index with no structural change but is systematically higher (Figure 6).
2. The accuracy of the estimation of the *OPHM* index is described in Figure 7. As for the index with no structural change, the explanatory power shown by the coefficient of determination is weaker in 1991–1992 (Figure 5). This is also the case for the Overlapping Period Hedonic Model index since the power for those periods that include 1991–1992 in the overlapped estimate period is weaker.

3. We can see the trend in this index more easily than in the index with no structural change. When we observe the changes in regression coefficient historically (Figures 8–11), they have moved more smoothly without violent fluctuation.
4. The regression coefficients for "distance from nearest station" and "proximity to CBD" become smaller in recent years in absolute terms. The elasticity of the distance becomes smaller. There is cyclical change for "age of building". The elasticity of "occupied area" becomes greater in recent years. In summary, consumer preference is moving towards "occupied area" rather than location.
5. Table 4 shows the average and standard deviation of regression coefficients for the *NRHI* and *OPHM*. Coefficients of variation (CV) in *NRHI* and *OPHM* model are: CV for nearest station is *NRHI* = –0.191, *OPHM* = –0.119, CV for proximity to CBD is *NRHI* = –0.381 , *OPHM* = –0.320, CV for age of building is *NRHI* = –0.209 , *OPHM* = –0.178 and CV for occupied area is *NRHI* = 4.277 , *OPHM* = 3.898. All these figures of *OPHM* are smaller than those of the model *NRHI*.

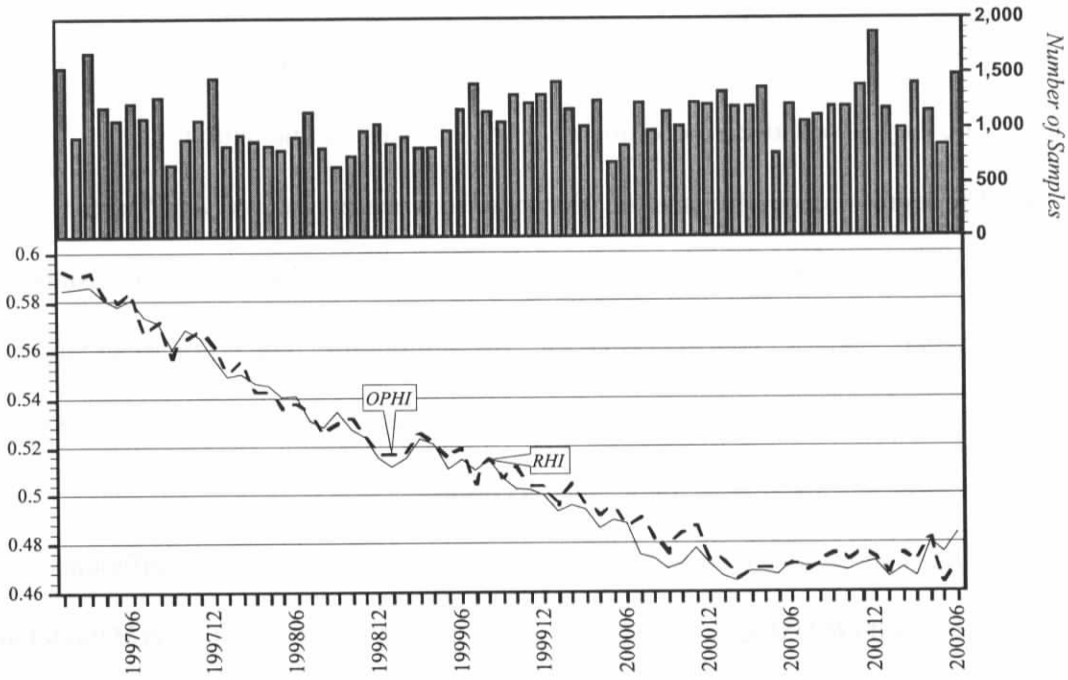

Figure 6. Comparison of price indices between *OPHM* and *RHI*: 1997/01 - 2002/06.

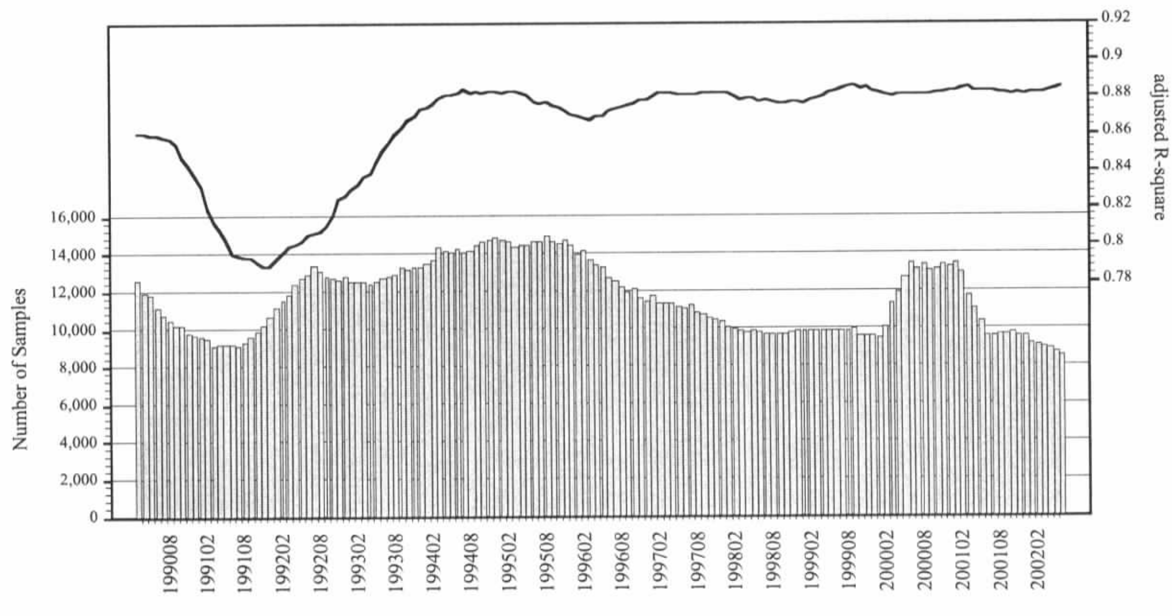

Figure 7. Accuracy of *OPHM*: 1989/04 - 2002/06

Table 4. Statistics of the main regression coefficients with NRHI and OPHM

Types of model	Principal Independent Variables	Summary statistics of estimated parameter				
		Average	Standard deviation	Coefficient of variation	Skewness	Kurtosis
NRHI	*WK*: Distance to the nearest station	-0.055	0.011	-0.191	-0.274	-0.216
	ACC: Accessibility to CBD	-0.406	0.155	-0.381	-0.619	-0.139
	BY: Age of building	-0.175	0.037	-0.209	0.231	-0.935
	FS: Floor space	0.024	0.101	4.277	-0.773	-0.881
OPHI	*WK*: Distance to nearest station	-0.056	0.007	-0.119	-0.317	-0.921
	ACC: Accessibility to CBD	-0.415	0.133	-0.320	-0.758	-0.536
	BY: Age of building	-0.170	0.030	-0.178	0.304	-1.289
	FS: Floor space	0.018	0.070	3.898	-0.645	-1.309

1990.03 - 2002.06:Monthly Number of Mode=149

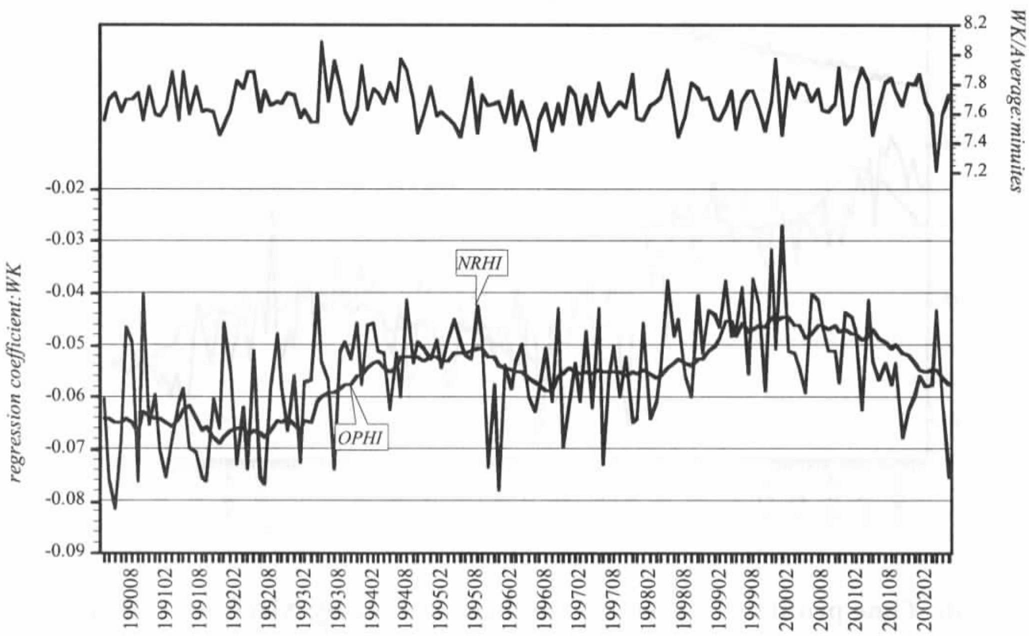

Figure 8. Time profile of regression coefficient of *WK* by NRHI and OPHI method

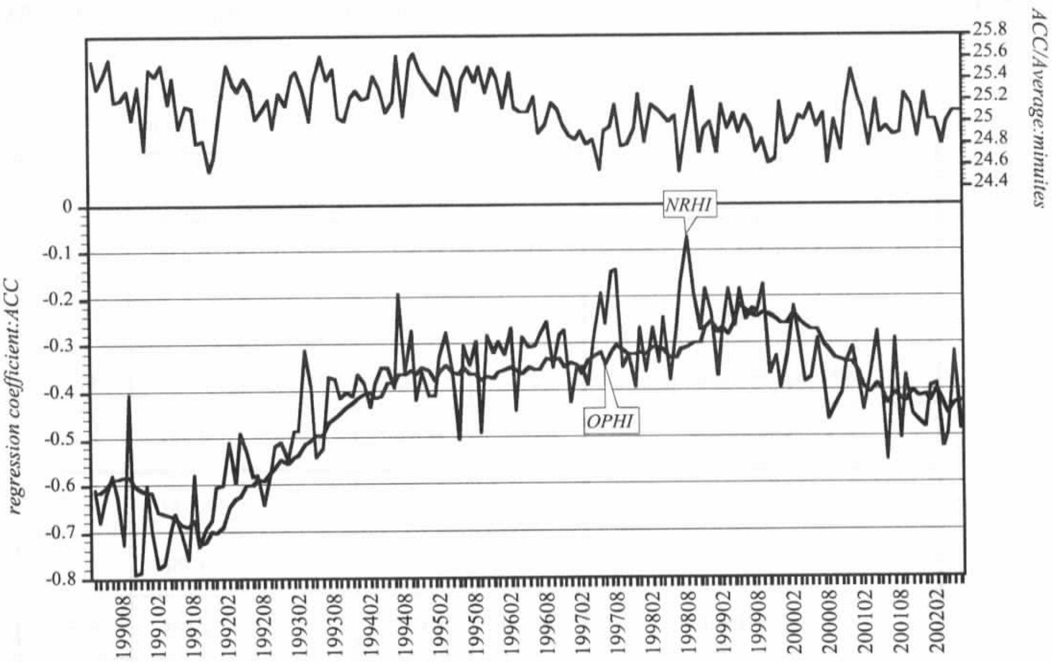

Figure 9. Time profile of regression coefficient of *ACC* by NRHI and OPHI method

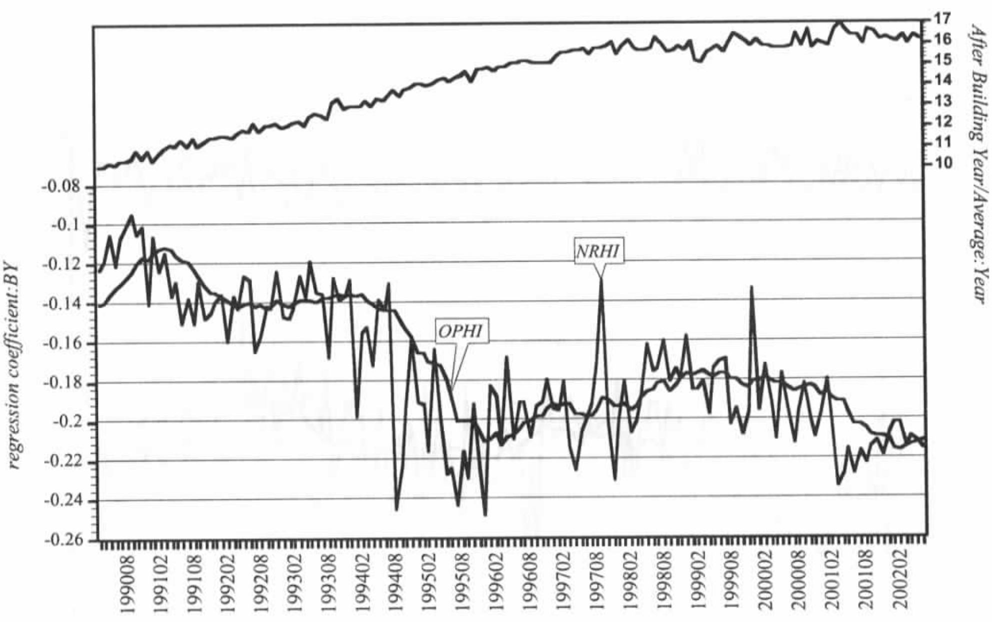

Figure10. Time profile of regression coefficient of *BY* by NRHI and OPHI method

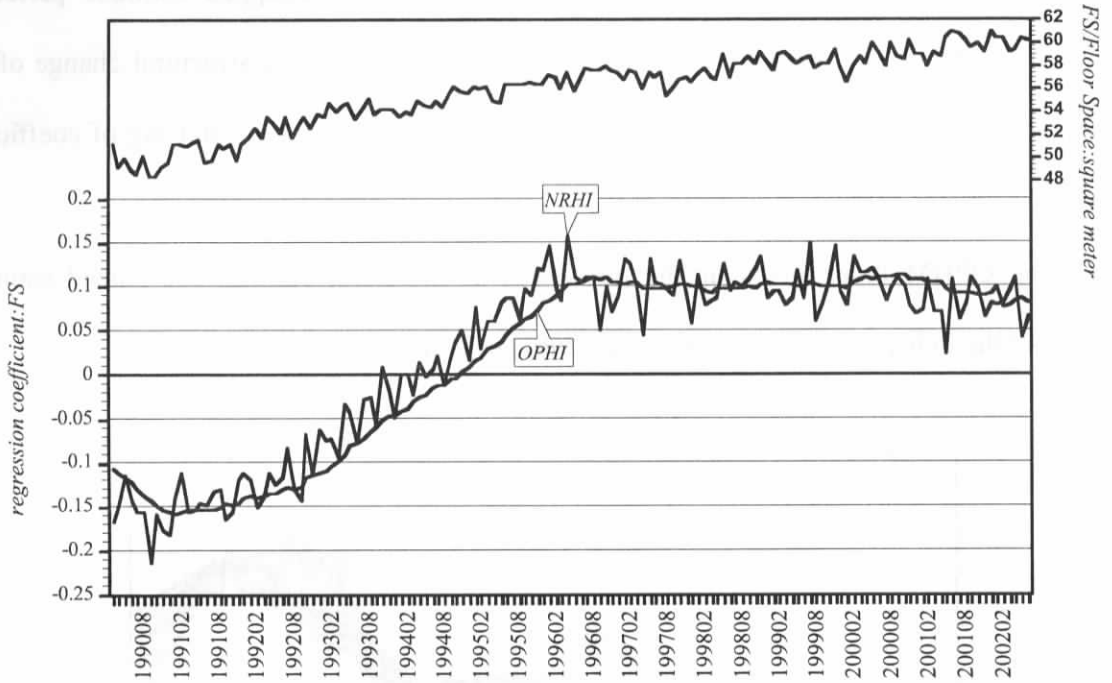

Figure11. Time profile of regression coefficient of *FS* by NRHI and OPHI method

5.3 Conclusion Regression coefficient variation by changing τ

The index *OPHI* based on the Overlapping Period Hedonic Model that we propose is more stable than *NRPI.* However, changing the overlapped estimate period τ gives us information of how sensitively each regression coefficient depends on τ. The following two points are to be settled.

- If an overlapped estimate period τ brings time lag structure into regression coefficients, the expected lag structure of the price index on the regression model must be specified.
- If the lag of an index originating in the overlapped period τ can be observed, standards and algorithms for optimizing the interval τ need to be developed.

Changing the period τ from 12 to 36 months leaves 25 temporal paths of regression coefficients. The paths of four main variables that explain variation in secondhand housing prices significantly are shown in Figures 12–15. Seemingly all four variables indicate evolving lag patterns in regression coefficients according to increases in overlapped estimate period τ. Regression coefficients show systematic fluctuation with time; i.e., the structural change of the hedonic model. Moreover, a wider estimate period τ corresponds to a greater lag of coefficient change and a greater delay in movement of the index up and down.

Thus, the OPHM index gives us the possibility of choosing optimal overlapped estimate period τ, and the index is expected to be either stable or dynamic.

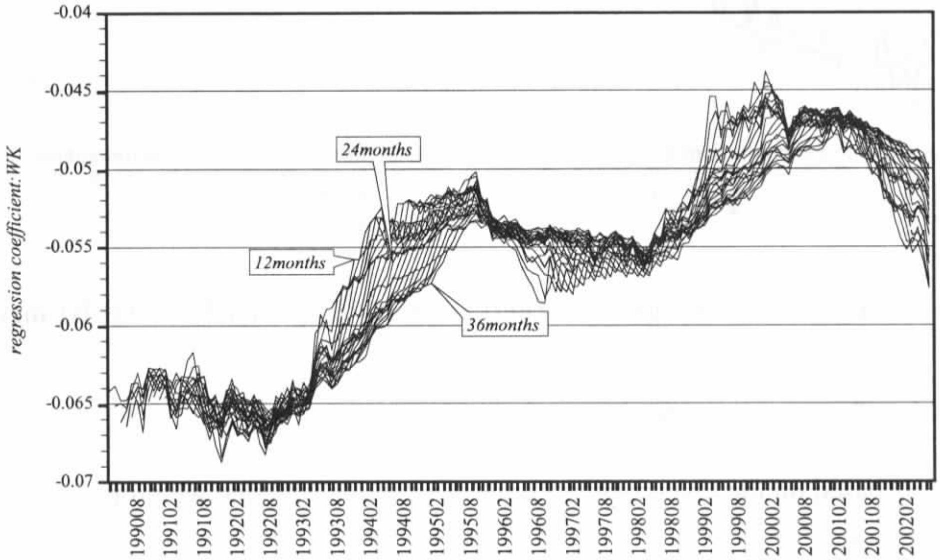

Figure12. Time profile of regression coefficient of *WK* with τ of 12 to 36 months

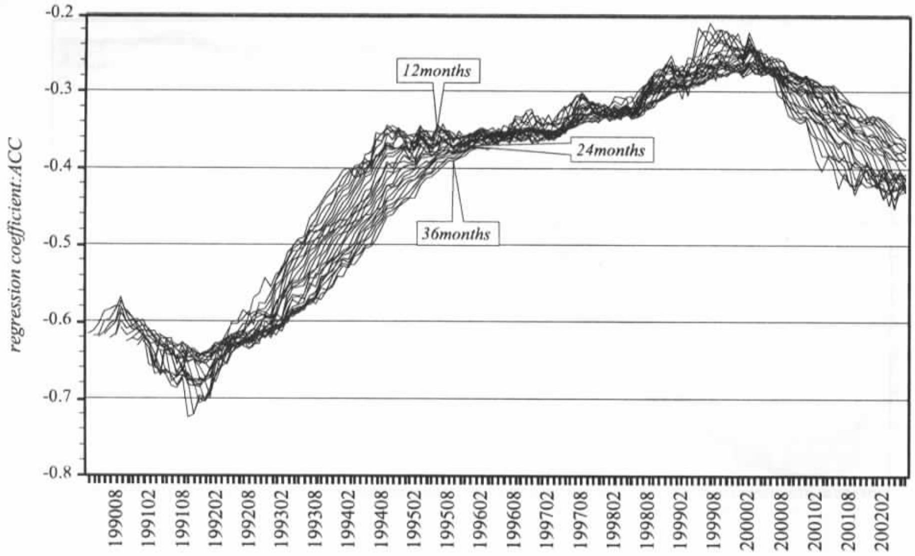

Figure13. Time profile of regression coefficient of *ACC* with τ of 12 to 36 months

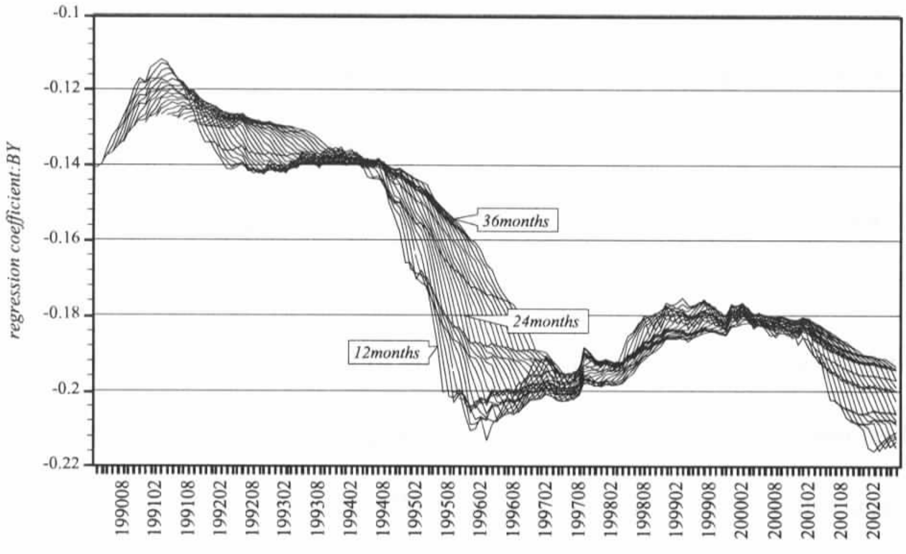

Figure14. Time profile of regression coefficient of *BY* with τ of 12 to 36 months

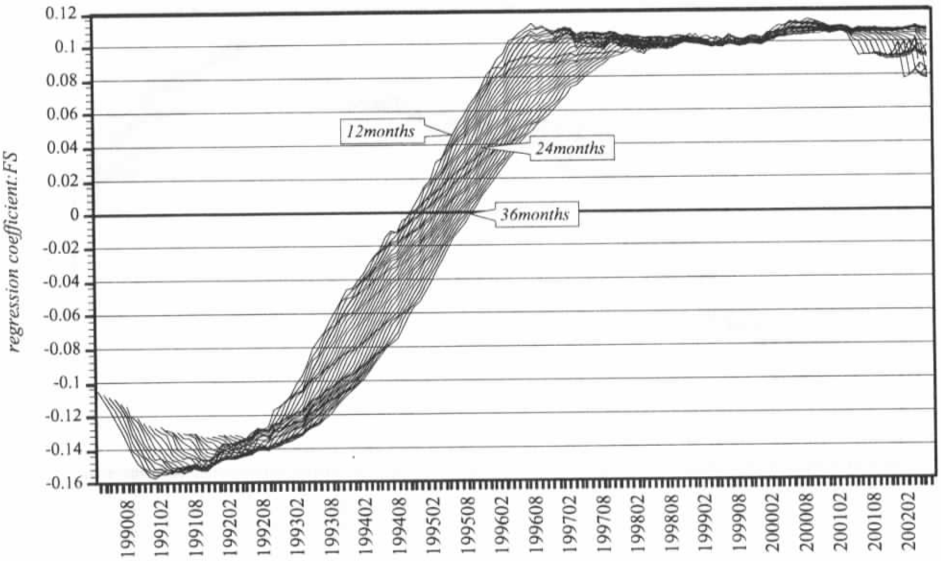

Figure15. Time profile of regression coefficient of *FS* with τ of 12 to 36 months

6 Conclusion

In the presumption of a hedonic index, the correspondence to structural change is a very important problem. It is understood that in an *NRHI*, correspondence to structural change is quite unstable. We propose *OPHM* as a conjunctive in presumption of a hedonic index with structural changes.

However, after the analysis of the *OPHM* index, the following issues remain.

1. The first issue is how to decide the overlapped estimate period. Our observation of both the index with no structural change and *OPHM* shows that regression coefficients change from period to period and that there are structural changes. We should not set the overlapped estimate period too long, as this may miss the structural change. At the same time, it should not be too short since both the regression coefficients and the price index become unstable. Although we changed the overlapped period from 12 months to

36 months and tried observation of change of a regression coefficient, when we proposed the optimal period, we did not obtain the desired results. We need clearer criteria. [9]

2. Secondly, there are structural differences between regions. For example, we understand that consumer preference is different between the east side and the west side of Tokyo's 23 wards. Also, we talk about the nature of locality and the status of location in our industry. If each region has a different structure, how can we have regional segments with the same structure? What is the relationship between the model consisting of Tokyo's 23 wards as a whole and individual models for each prefecture in the region? Additionally, we have not analyzed in this paper whether the rate of structural change has been the same historically and regionally.
3. Thirdly, the number of observations is different in each month if the sample is divided on a monthly and regional basis. In panel data analysis, it is known that the number of observations varies either if the observable market information is lopsided with regard to particular characteristics or if observations of particular characteristics are missed. [10] The issues are how to deal with these biases, whether we can avoid the problem by using a regional dummy factor and what effects are given by the biases in the identification of differences in regional structure.
4. Fourthly, seasonal change of the number of observations of a particular set of information at a particular moment greatly affects estimation of regression coefficient of the index with structural change and the *OPHM* index. Without structural change, the

9 We checked that structural change does not happen in less than three years on our data.

10 As for sampling selection bias, see Griliches (1996), Davidson and Mackinnon (1993) and Maddala (1985).

estimate results would be unbiased since the observations are drawn from the same sample while they are biased if there are any structural changes.

We have proposed the *OPHM* index since neither the index with no structural change nor the index with structural change is adequate for coping with structural change. Several tasks remain to be resolved, but the proposed index overcomes the defects of both the index with no structural change and the index with structural change.

The identified tasks are summarized as follows.

1. The first task is how to resolve the nonlinear structure of a hedonic price index. In this paper, we undertake nonlinear analysis of the relationship between condominium price and variables such as the distance from the nearest station by logarithmic transformation. However, the relationship should be considered a nonlinear structure. Several previous studies carry out nonlinear estimation by using Box–Cox transformation. We need to develop a more suitable function.
2. The second task is to sort out heterogeneous variance in the monthly database. If we can confirm that the variance is constant between each period, we can work out the problem by the least squares method, weighted by estimated variance for each period. However, we have no proof of this. It may be worth investigating the application of the generalized least squares method.

References

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press.

Allen, R. G. D. (1975). *Index Numbers in Theory and Practice*, The Macmillan Press.

Bowles, G., P. McAllister, and H. Tarbert (2001). "An Assessment of the Impact of Valuation Error on Property Investment Performance Measurement," *Journal of Property &Finance* 19(2), 139–155.

Clayton, J. (1996). "Rational Expectation, Market Fundamentals and Housing Price Volatility," *Real Estate Economics*, 24(4), 441–470.

Davidson, R., and Mackinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

Garcia, A. R., and P. Perron (1996). "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics*. 78, 111–125.

Geltner, D. (1997). "The Use of Appraisal in Portfolio Valuation and Index Construction," *Journal of Property Valuation and Investment*, 15(5), 423–447.

Griliches, Z. (1996). "Introduction to the Application." In: L. Matayas and P. Sevestre, eds., *The Econometrics of Panel Data*. Kluwer Academic Publishers, pp. 655–659.

Harvey, A. C. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hidano, N. (2003). *The Economic Valuation of the Environment and Public Policy: A Hedonic Approach*. Edward Elgar Publication.

Jushan, B., and P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66(1), 47–78.

Maddala, G. S. (1983). "Limited-dependent and Qualitative Variables," *Econometrics*. Cambridge University Press.

Maddala, G. S. and I. M. Kim (1998). *Unit Roots Cointegration and Structural Change*, Cambridge University Press.

Nishimura, K., and C. Shimizu (2003). "Distortion in Land Price Information—Mechanism in Sales Comparables and Appraisal Value Relation," Discussion Paper CF195, Center for International Research on the Japanese Economy, Faculty of Economics, University of Tokyo. Retrieved Feb. 2003 from http://www.e.u-tokyo.ac.jp/cirje/research/dp/2003/list.htm

Ono, H., H. Takatsuji, and C. Shimizu (2002). "Estimation of Residential Price Model of Second-hand Condominium based on the Quality," *RIPESS Working Paper No.1*, Reitaku University.

Ono, H., H. Takatsuji, and C. Shimizu (2003). "Estimation of Hedonic Price Model for Secondhand Housing Market (in Japanese: Hinshitsu wo Kouryoshita Chuuko Manshon Kakaku Moderu no Suitei)," Reitaku International Journal of Economic Studies, 10(2), 81–102.

Shepard, S. (1999). "Hedonic Analysis of Housing Markets." In: Paul Cheshire and Edwin S. Mills, eds., *Handbook of Regional and Urban Economics*, 3. Elsevier Science Ltd., pp. 1595–1635.

Takatsuji, H., H. Ono, and C. Shimizu (2003). "Conjunct Hedonic Price Index in the Secondhand Housing Market with Structural Changes (in Japanese: Kouzouhenka no aru Kakakukansuu wo mochiita Hinshitu Chouseizumi Juutaku Kakakushisuu no Setsuzokuhou)," Reitaku International Journal of Economic Studies, 10(2), 103–134.