Grant-in-Aid for Scientific Research（S）

Real Estate Markets, Financial Crisis, and Economic Growth

: An Integrated Economic Approach

Working Paper Series No.34

Property Price Index Theory and Estimation : A Survey

Chihiro Shimizu

and

Koji Karato

January, 2016

HIT-REFINED PROJECT

Institute of Economic Research, Hitotsubashi University

Naka 2-1, Kunitachi-city, Tokyo 186-8603, JAPAN

Tel: +81-42-580-9145

E-mail: hit-refined-sec@ier.hit-u.ac.jp

http://www.ier.hit-u.ac.jp/hit-refined/

Property Price Index Theory and Estimation: A Survey*

Chihiro Shimizu† and Koji Karato‡

February 12, 2016

Abstract

Property has the particularity of being a non-homogeneous good, and based on this, it is necessary to perform quality adjustment when estimating property price indexes. Various methods of quality adjustment have been proposed and applied, such as the hedonic method often used in price statistics and, due to the fact that the information that can be used in estimation is limited, the repeat sales price method, methods using property appraisal price information, and so forth. However, since there is a lack of theoretical knowledge and data restrictions, it is no exaggeration to say that it is difficult to evaluate their practical application in the present situation. Therefore, focusing on the hedonic method that has been proposed as a quality adjustment method for property price indexes, in addition to repeat sales price indexes and indexes employing property appraisal prices, this paper aimed to outline the underlying econometric theory and clarify the advantages and disadvantages of the respective estimation methods.

Key Words:
Hedonic price index; Repeat sales price index; Age effect; Hybrid method; Property appraisal price method; SPAR

Journal of Economic Literature Classification Numbers:
C2, C23, C43, D12, E31, R21.

* This research was conducted by the authors while participating in the Hitotsubashi Project on Real Estate, Financial Crisis, and Economic Dynamics, (HIT-REFINED) supported by a JSPS Grant-in-Aid for Scientific Research (S). The authors thank Erwin Diewert, David Geltner and Marc Francke for helpful discussions.

† Professor, Institute of Real Estate Studies, National University of Singapore, cshimizu@nus.edu.sg.
‡ Professor, Faculty of Economics, University of Toyama. kkarato@eco.u-toyama.ac.jp.

1 Introduction

The formation and collapse of property bubbles has a profound impact on the economic administration of many leading nations. The property bubble that began around the mid-1980s in Japan has been called the 20th century's biggest bubble. In its aftermath, the country faced a period of long-term economic stagnation dubbed the "lost decade." Many countries had similar experiences with this kind of problem – for example, Sweden's economic crisis in the 1990s and the global financial crisis and economic stagnation caused by the formation and collapse of the U.S.-centered property bubble in the early 21st century.

In light of this, it was pointed out that the "information gap" which existed between policy-making authorities and the property (including housing) and financial markets was a problem. In 2009, the IMF proposed the creation of a housing price index to the G20 in order to fill in this information gap, and the proposal was adopted. Furthermore, in 2011, it was suggested that the next economic crisis would be caused by land – i.e., profit-generating property (commercial property) – and it was decided to create a commercial property index as well. But how should these property price indexes be created?

Property standards and facilities vary to a greater or lesser extent for each building, so there is no such thing as identical properties. Even if one assumed that the standards and facilities were the same, the process by which quality deteriorated would differ by building age, so the buildings would become non-homogeneous over time. In other words, property has the particularity of being a non-homogeneous good. In addition to this problem, the development of building technology is relatively fast, so quality changes over time. That is, not only does a building's functionality decline over time, but it becomes economically obsolescent with the advance of technology. As well, in cases where the surrounding environment changes significantly through redevelopment and the like, location characteristics such as transport accessibility of the city center also change.

When attempting to capture temporal fluctuations in property prices while dealing with the problems caused by property being a non-homogeneous good and changes in quality, it is necessary to perform quality adjustment.

In order to address these problems, there are quite a few points that can be adapted from existing index theory, as typified by consumer price statistics. For example, with regard to changes in quality accompanying technological development, the quality adjustment method known as the hedonic approach is used in private vehicle price statistics and the like. It would therefore be natural to also consider quality adjustment with the hedonic method for property price indexes, since this enables consistency with other types of economic statistics.

However, when it comes to methods of quality adjustment for property price indexes, if one looks at the *Residential Property Price Indices Handbook* published by EuroStat in 2013, it present a variety of methods along with their advantages and disadvantages: a) Stratification or Mix Adjustment Methods, b) Hedonic Regression Methods, c) Repeat Sales Methods, and d) Appraisal-Based Methods. This is because, in reality, multiple methods have been applied in the estimation of property price indexes.

Why have approaches other than the hedonic method been applied in practice?

The first reason is the difficulty of quality adjustment. As explained previously, the reason for performing quality adjustment of property is that it is a good for which no homogeneity exists, so it is strongly heterogeneous. In such a case, in addition to the problems relating to quality changes faced in consumer price statistics and the like, one must also address said heterogeneity. In other words, quality adjustment involves a high degree of difficulty.

The second reason is the lack of usable price information at the micro level when estimating property price indexes. If attempting to apply the hedonic method, not only transaction price, transaction time, and land/building size but also location-related information such as the time to the city center and detailed information related to the building age and features are required. When there is no such information, the price index must be estimated with limited data. With the repeat sales method, quality adjustment is possible with just the transaction price and transaction time, so it has the advantage of minimizing the information needed with respect to property-related variables. That being the case, when attempting to measure price changes when information is limited, creating an index using only data for properties that are transacted repeatedly is consistent with the general thinking behind price index estimation methods.

However, unlike other goods and service markets, the property transaction market is extremely thin. Thus, when attempting, for example, to create a monthly price index, one may easily anticipate that many problems will occur, since – unlike markets where goods and services of identical quality are transacted frequently – this is a market in which property with identical characteristics is transacted only once every few years.

Third, in actual property transaction practice, it is not uncommon for property appraisal prices to be used. Not only is property strongly heterogeneous but there are few transactions – depending on the region and usage, there are even markets where almost no transactions exist. In light of this, when trying to determine prices, using prices based on property appraisals is a valid approach.

Thus, attempting to estimate property price indexes involves many difficulties. Compared to the housing market, where there is a relatively large number of transactions and the quality gap is small (i.e., it is more homogeneous), these difficulties mean that for commercial property (offices, retail facilities, hotels, logistics facilities, hospitals, farmland, etc.): 1) there will be more heterogeneity and a greater lack of information, 2) there will be a reduction in repeat sales samples, and 3) there is a greater probability of property appraisal prices being used.

For markets for which property price information is relatively easy to obtain, the aim of this paper is to outline the characteristics of the hedonic method and repeat sales method that may be used in creating property price indexes, as well as estimation methods using property appraisal prices. Specifically, focusing on the hedonic method and repeat sales method, it will provide a comprehensive survey relating to quality adjustment methods when estimating property price indexes and clarify the characteristics of the various estimation methods. What's more, drawing on this outline, it will present a view of how property price indexes should be created, from the perspective of estimation method theory.

2 The Hedonic Price Method

2.1 The Hedonic Approach

The hedonic approach is a technique established theoretically by Rosen (1974)[55]. Specifically, it treats a given product's price as an aggregate (bundle of attributes) of the values of the product's various attributes (characteristics) and estimates the various attribute prices using regression analysis. For many products circulating on the market, even when their intended use is the same, considerable differentiation exists based on their performance, functions, etc. Differences in attributes are reflected in the product's production costs. One could also say that consumer evaluations of the product's specific performance and functions are also reflected in the price determined by the market. However, the attributes themselves are not necessarily bought and sold on the market. With the hedonic approach, by regressing product price on variables representing attribute quality and quantity, it is possible to measure the shadow price of non-market goods based on the estimated coefficient value.

Lancaster (1966)[44] has conducted theoretical analysis of consumer behavior based on the assumption that consumer utility depends not on the product itself but on the various features, functions, etc., that comprise the product. The product's market price is thought to be determined based on supply and demand in relation to its various characteristics. However, the market with respect to these characteristics is not necessarily explicit but may be hidden in the background of product price determination. Lancaster's aim was to explicitly treat such underlying mechanisms and analyze consumer behavior in differentiated goods markets.

Rigorously examining the relationship between differentiated product prices and consumer behavior is essential in preparing price indexes. For example, in the case of digital consumer electronics, passenger vehicles, housing, etc., even if the price is the same, quality will improve and functions increase as time passes. With the Laspeyres method, since a market basket is fixed at a baseline point in time, price indexes based on this method ignore changes in quality and functionality. Using the hedonic approach helps estimate the performance ratio between new and old products.

Rosen (1974)[55]'s price analysis of differentiated goods is a study that theoretically clarifies the manner in which product prices comprised by bundles of attributes are generated on the market. The study rigorously examines the relationship between the product supplier offer function, product demander bid function, and hedonic market price function, and characterizes the market price of products based on consumer and producer behavior. When this hedonic market price function is used, it is possible to obtain the acceptable payment amount for product attributes.

Section 2.2 below outlines Rosen (1974)[55]'s hedonic approach theory, while Section 2.3 addresses issues relating to the estimation of hedonic market price functions. Following Diewert, Heravi and Silver (2007)[27], Section 2.4 summarizes differences based on a hedonic dummy index and hedonic imputed index. Section 2.5 explains the characteristics of a producer price-related quality-adjusted hedonic index. Section 2.6 summarizes the characteristics of the hedonic price method.

2.2 Hedonic Approach Theory

2.2.1 The Bid Function

Following Rosen (1974)[55]'s method, we will demonstrate the theoretical basis of the hedonic approach, using real estate as an example. The value of characteristic k comprising real estate shall be expressed as $z_k \, (k = 1, 2, \dots, K)$. Real estate characteristics represent size, building structure, kitchen, bathroom, accessibility of transportation, natural environment, social environment, and so forth. According to Rosen, the relationship between real estate market price p and characteristic value $z_1, \dots, z_k, \dots, z_K$ may be expressed with the following hedonic price function h:

$$p = h(z_1, \dots, z_k, \dots, z_K) \tag{1}$$

The main objective of Rosen's analysis is to clarify how (1) is determined by the market.

Given market price function (1), consumers select the optimal combination of real estate characteristics. The issue of utility maximization may be formulated as follows:

$$\max_{x,\boldsymbol{z}} U(x, \boldsymbol{z}) \tag{2}$$

$$s.t.\ I = x + h(\boldsymbol{z}) \tag{3}$$

Here, $U(\cdot)$ is a well-behaved, strictly concave function, x is composite goods including goods and services other than real estate, $\boldsymbol{z} = (z_1, \ldots, z_k, \ldots, z_K)$ is the real estate characteristic vector, and I is income. The composite goods price is standardized as 1. Based on the parameters of this optimization issue step, $U_k/U_x = h_k(\boldsymbol{z})$ is established. Note that $U_k = \frac{\partial U(x,\boldsymbol{z})}{\partial z_k}$, $U_x = \frac{\partial U(x,\boldsymbol{z})}{\partial x}$, and $h_k(\boldsymbol{z}) = \frac{\partial h(\boldsymbol{z})}{\partial z_k}$. In other words, this shows that the marginal utility of the real estate characteristic measured using the marginal utility of income is equal to the marginal contribution value of the attribute in market prices.

It is possible to determine the market price function using the bid function. Based on a given utility level u and income I, if the bid offered by a housing demander for real estate possessing characteristic $\boldsymbol{z}$ is taken as θ, then based on (2), this may be written as $U(I - \theta, \boldsymbol{z}) = u$.
If one solves this for θ, the amount that a consumer is able to spend on housing with respect to various combinations of characteristic $\boldsymbol{z}$ may be expressed as the bid function $\theta(\boldsymbol{z}; I, u)$, given the utility level and income. In order to raise (lower) the utility level u, the bid for housing with characteristic $\boldsymbol{z}$ must decrease (increase) $\left(\frac{\partial \theta(\boldsymbol{z};I,u)}{\partial u} = -U_x^{-1} < 0\right)$. Therefore, this shows that θ, when it reaches utility level u, is the maximum price that may be paid for housing with characteristic $\boldsymbol{z}$.

Based on (2), (3), and the bid function $\theta(\boldsymbol{z}; I, u)$, one may write that $U(I - \theta(\boldsymbol{z}; I, u), \boldsymbol{z}) = u$. If this formula is partially differentiated for z_k and 0 is included, the following is obtained:

$$-U_x \frac{\partial \theta(\boldsymbol{z}; I, u)}{\partial z_k} + U_k = 0$$

When the utility is maximized at the level of u^*, since $U_k/U_x = h_k(\boldsymbol{z}^*)$ for the optimal combination of characteristics $\boldsymbol{z}^*$, the following two equations are definitely established:

$$\frac{\partial \theta(\boldsymbol{z}^*; I, u^*)}{\partial z_k} = h_k(\boldsymbol{z}^*) \tag{4}$$

$$\theta(\boldsymbol{z}^*; I, u^*) = h(\boldsymbol{z}^*) \tag{5}$$

(4) and (5) show that when the optimal characteristics are selected, the slope of the bid function and the slope of the market price function are consistent and the bid and market price are also equal. In other words, based on the optimal characteristic value, the bid function and market price function are contiguous.

When consumer incomes and preferences vary, the bid function also varies. However, since the bid function and market price function must be contiguous in market equilibrium, the market price function is an envelope of the bid function for all consumers, with their various incomes and preferences.

2.2.2 The Offer Function

It is also possible to define the price offer function for real estate suppliers and theorize the relationship with the market price function from the issue of profit maximization. For a given level of technology, the offer function is the minimum price offered when a given profit is reached. When a company selects the optimal characteristics and produces real estate, the slope of the offer function and the slope of the market price (per unit of real estate) function will be consistent based on profit maximization behavior, and the offer price and market price will also be consistent. Therefore, based on the optimal characteristic value, the offer function and market price function are contiguous. Since heterogeneity exists in real estate producers' technology, offer prices also vary in accordance with this. Since the offer price and market price need to be consistent in equilibrium, the market price function is an envelope of the offer function for various companies.

Based on the above, the hedonic market price function is an envelope of both the bid function for an infinite number of real estate demanders and the offer function for an infinite number of real estate suppliers. As well, in the case of there being one supplier company, the bid function is equal to the marginal cost if one additional unit of real estate is produced (or the average cost per unit of real estate). As a result, the market price function is equal to the supplier's marginal cost.

2.2.3 Willingness to Pay

If the bid function is used, it is possible to obtain consumers' willingness to pay with respect to changes in attributes. For $\boldsymbol{z}^*$, let us now assume that $p^* = \theta(\boldsymbol{z}^*; I, u^*) = h(\boldsymbol{z}^*)$. When real estate K's characteristic z_K^* is increased to (z_K^{**}), the demander's willingness to pay (WTP) may be defined with the following formula:

$$WTP \equiv \theta\left(z_1^*, \ldots, z_{K-1}^*, z_K^{**}; I, u^*\right) - p^* \tag{6}$$

In other words, when the characteristic value changes incrementally, the willingness to pay is the additional value that may be paid for real estate without changing the utility level. Since the utility function U is

$$\frac{\partial^2}{\partial z_k^2}\theta(\boldsymbol{z}^*; I, u^*) = \frac{U_x^2 U_{kk} - 2U_x U_k U_{xk} + U_k^2 U_{xx}}{U_x^3} < 0$$

when it is a strict concave function (the Hessian matrix is a negative definite matrix), the bid function is a concave function. (4) and (5) are established based on the optimal characteristic value combination $\boldsymbol{z}^*$, and given that the bid function is a concave function, one can derive from

$$\theta\left(z_1^*, \ldots, z_{K-1}^*, z_K^{**}; I, u^*\right) < h\left(z_1^*, \ldots, z_{K-1}^*, z_K^{**}\right) = p'$$

That

$$p' - p^* > WTP \tag{7}$$

In other words, caution is required with regard to the limit value of market price function characteristics, since as long as demanders are not homogeneous, it is possible that the willingness to pay is overestimated. However, if it is assumed that changes in characteristic values will be sufficiently small, the market price function limit value may be used as an approximation of the willingness to pay.

2.3 Hedonic Market Price Function Estimation

2.3.1 Function Types

In order to accurately measure willingness to pay, estimation of the bid function is required, but in general an approximation is used by estimating the hedonic market price function (1).

When estimating the hedonic market price function, the function type is an issue. Given that simple estimation is possible, models such as double logarithms, semi-logs, and line shapes are often used.

When real estate price at multiple points in time is observed as data, the hedonic market price at time t for property n may be described with the following formula:

$$y_n^t = \alpha_t + \boldsymbol{z}_n^{t\prime}\boldsymbol{\gamma} + \varepsilon_n^t \quad (n = 1, 2, \ldots, N(t);\ t = 0, 1, \ldots, T) \tag{8}$$

Here, y_n^t is the housing price logarithm ($\ln p_n^t$) or exact numeric value (p_n^t), α_t is the unknown time effect, $\boldsymbol{z}_n^t = (1, z_{n1}^t, \ldots, z_{nk}^t, \ldots, z_{nK}^t)'$ is the explanatory variable (characteristic) vector including a constant term, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_k, \ldots, \gamma_K)$ is the coefficient vector, and ε_n^t is the error term. As an example, a semi-log model including the time effect may be written as:

$$y_n^t = \ln p_n^t = \alpha_t + \gamma_0 + \sum_{k=1}^{K} \gamma_k z_{nk}^t + \varepsilon_n^t \tag{9}$$

In this model, the estimation value of coefficient $\boldsymbol{\gamma}$ shows the effect of the characteristic value with respect to real estate price, and if a dummy variable is used for each point in time, estimation may be made estimated based on the method of least squares. To avoid multicollinearity and distinguish all parameters, it is necessary to perform some kind of standardization for α_t and γ_0. Typically, at the observation starting point $t = 0$, it is considered that $\alpha_0 = 0$, and a dummy variable for each point in time is used with respect to $t = 1, 2, \ldots, T$.

Since the function type of the hedonic market price function h cannot be specified in theoretical terms, it must be selected with a statistical technique. Even if specified in a double logarithmic model or semi-log model, the form is not necessarily the ideal one. Studies from the 1980s onward, such as Linneman (1980)[45], have performed non-linear estimation using Box-Cox transformation. In this case, the left side of (9) can be rewritten as follows:

$$y = \begin{cases} \dfrac{p^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln p & \lambda = 0 \end{cases} \tag{10}$$

Here, λ is an unknown parameter. Halvorson and Polakowski (1981)[36] tested various function forms by applying Box-Cox transformation to a flexible function form using a two-step approximation formula including a cross-term between explanatory variables. In response to their paper, Cassel and Mendelsohn (1985)[11] increased the explanatory power by including multiple cross-terms between variables, but pointed out that there is a reduction in the reliability of the coefficient estimation value due to multicollinearity and that interpretation of the marginal effect of hedonic characteristic values becomes more difficult. Cropper, Deck, and McConnell (1988)[13] performed statistical tests based on a translog form and Diewert-type utility function (Barten (1964)[3], Diewert (1971)[19], Diewert (1973)[20]), showing that if observational errors are included in the variables, a linear model or linear Box-Cox transformation model is superior to quadratic form Box-Cox transformation when it comes to formulation.

There is also research that has proposed using a non-parametric method or semi-parametric method instead of a parametric function form to formulate the hedonic price function. With these approaches, attribute prices are inferred directly from the data without specifying a function form in advance (Knight et al. 1993[40], Anglin and Gencay (1996)[1], Pace 1995[53]). However, it has also been pointed out that, as with parametric analysis techniques, these do not free one from data-related problems (multicollinearity). In tests relating to the selection of parametric versus non-parametric models, Anglin and Gencay (1996)[1] have shown that it is relatively easy to dismiss parametric models. It is not that the parametric model variable structure is weak; rather, this result was demonstrated even for parametric models that passed a number of standard tests for model selection. Using a more flexible Generalized Additive Model (GAM), Pace (1998)[52] estimated a semi-parametric-type hedonic price function and demonstrated that it was superior to all types of parametric model. Since GAM itself is an established statistical technique, this finding shows that the incorporation of a non-parametric method in the hedonic approach is extremely effective.

2.3.2 The Problem of Distinguishing the Marginal Bid Value Function

If characteristic values have a significant effect on market prices, the willingness to pay will cause divergence between the hedonic market price function and bid value function, so it is necessary to estimate the bid value function or bid value marginal effect. As a method of estimating the bid value function, Rosen (1974)[55] has proposed a method that regresses the market price function marginal effect on characteristic values and other exogenous variables.

$$\hat{h}_k = D_k(\boldsymbol{z}, \boldsymbol{A}) \tag{11}$$

$$\hat{h}_k = S_k(\boldsymbol{z}, \boldsymbol{B}) \tag{12}$$

Here, $\hat{h}_k$ is the marginal effect for hedonic market price function characteristic k, $D(\cdot)$ and $S(\cdot)$ are the characteristic's demand and supply functions, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are vectors showing the real estate demander and supplier type, respectively (based on income, manufacturing technology, etc.). Since the marginal effect is the shadow price of the characteristic value, (11) and (12) are supply and demand simultaneous equations using inverse demand (bid value) and inverse supply (offer price), and supply and demand are distinguished using $\boldsymbol{A}$ and $\boldsymbol{B}$ as instrumental variables.

Following Rosen's model, Witte, Sumka, and Erekson (1979)[**?**] estimated simultaneous equations for three characteristics covering multiple housing markets. However, as Brown and Rosen (1982)[6] have pointed out, it is not possible to properly distinguish between characteristic value supply and demand with estimation based on this method. Since the market price function marginal effect $\hat{h}_k$ estimated in the first step is derived from $h(\boldsymbol{z})$, one may consider that the characteristic price shown with the marginal effect is also a function of $\boldsymbol{z}$. Demand for $\boldsymbol{z}$ depends on the various characteristic prices, and there is a correlation between characteristic prices and characteristic demand equation errors. In other words, it is possible that the effect of characteristic prices on characteristic demand is estimated with a bias.

This problem of distinguishing the bid value function and offer function has been considered by Diamond and Smith (1985)[18] and Mendelsohn (1985)[49]. First, with regard to estimation of the first step hedonic market price function, it is pointed out that, apart from characteristic vectors, there is a need for exogenous variables not included in either the bid value function or offer function as well as for a characteristic value exponential term. Then, in the second step, a marginal bid value function simultaneous equation system is estimated simultaneously using exogenous variables solely to meet the distinction conditions. Sheppard (1999)[56] has discussed the distinction problem in greater detail. Ekeland, Heckman, and Nesheim (2004)[29] and Heckman, Matzkin, and Nesheim (2010)[39] proposed a distinction method for hedonic price estimation using a non-parametric approach.

2.4 Price Index Estimation Based on the Hedonic Approach

2.4.1 Time Dummy Hedonic Regression

The hedonic approach is a useful technique when creating quality-adjusted price indexes. There are two representative types of hedonic price index: (i) time dummy hedonic indexes and (ii) imputed hedonic indexes. Following Diewert, Heravi and Silver (2007)[27], we discuss differences between the two types of price index below.

In (8), taking the observation period as two points in time, $(t = 0, 1)$, one may assume the following estimation model that regresses logarithmic price on an explanatory variable vector with the time dummy and constant term excluded:

$$y_n^t \equiv \ln p_n^t = \alpha_t + \sum_{k=1}^{K} \gamma_k z_{nk}^t + \varepsilon_n^t \quad (n = 1, 2, \ldots, N(t);\ t = 0, 1) \tag{13}$$

Here, α_t shows the average level of the product's quality-constant price for each period, and the overall scale of logarithmic price changes from time 0 to time 1 is $\alpha_1 - \alpha_0$.

Let us take $\mathbf{1}_t$ as an $N(t)$ dimension vector comprising everything from 1 and $\mathbf{0}_t$ as an $N(t)$ dimension vector comprising everything from 0. As well, let us take $\boldsymbol{y}^0$ and $\boldsymbol{y}^1$ as the $N(0)$ and $N(1)$ dimension vectors for the time 0 and time 1 logarithmic prices respectively, $\mathbf{Z}^0$ and $\mathbf{Z}^1$ as the $N(t) \times K$ explanatory variable matrices for time 0 and time 1 respectively, and $\boldsymbol{\varepsilon}^0$ and $\boldsymbol{\varepsilon}^1$ as the $N(0), N(1)$ dimension error vectors for time 0 and time 1 respectively. If we represent (13) as matrices for time 0 and time 1, they may be written as follows:

$$\boldsymbol{y}^0 = \mathbf{1}_0\alpha_0 + \mathbf{0}_0\alpha_1 + \mathbf{Z}^0\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^0 \tag{14}$$

$$\boldsymbol{y}^1 = \mathbf{0}_1\alpha_0 + \mathbf{1}_1\alpha_1 + \mathbf{Z}^1\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^1 \tag{15}$$

Here, if we take $\alpha_t^*, \boldsymbol{\gamma}^*$ as estimators based on the method of least squares, one can formulate the following using the estimators and the realized value $\boldsymbol{e}^t$ of the least squares residual error:

$$\boldsymbol{y}^0 = \mathbf{1}_0\alpha_0^* + \mathbf{0}_0\alpha_1^* + \mathbf{Z}^0\boldsymbol{\gamma}^* + \boldsymbol{e}^0 \tag{16}$$

$$\boldsymbol{y}^1 = \mathbf{0}_1\alpha_0^* + \mathbf{1}_1\alpha_1^* + \mathbf{Z}^1\boldsymbol{\gamma}^* + \boldsymbol{e}^1 \tag{17}$$

For (16) and (17), if we define $\boldsymbol{y} = \begin{bmatrix}\boldsymbol{y}^{0\prime} & \boldsymbol{y}^{1\prime}\end{bmatrix}'$ $((N(0) + N(1)) \times 1$ vector), $\boldsymbol{e} = \begin{bmatrix}\boldsymbol{e}^{0\prime} & \boldsymbol{e}^{1\prime}\end{bmatrix}'$ $((N(0)+N(1)) \times 1$ vector), $\boldsymbol{\varphi}^* = \begin{bmatrix}\alpha_0^* & \alpha_1^* & \boldsymbol{\gamma}^{*\prime}\end{bmatrix}'$ $((2+K) \times 1$ vector), and $\mathbf{X} = \begin{bmatrix}\mathbf{1}_0 & \mathbf{0}_0 & \mathbf{Z}^0 \\ \mathbf{0}_1 & \mathbf{1}_1 & \mathbf{Z}^1\end{bmatrix}$ $((N(0) + N(1)) \times (2 + K)$ matrix), then (16) and (17) may be rewritten as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\varphi}^* + \boldsymbol{e} \tag{18}$$

Here, since $\mathbf{X}$ and the residual error $\boldsymbol{e}$ are orthogonal, we can obtain:

$$\mathbf{X}'\boldsymbol{e} = \mathbf{X}'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\varphi}^*) = \mathbf{0}_{2+K} \tag{19}$$

In other words, we can obtain $\mathbf{1}_0'\boldsymbol{e}^0 = 0, \mathbf{1}_1'\boldsymbol{e}^1 = 0$ and $\mathbf{Z}^{0\prime}\boldsymbol{e}^0 + \mathbf{Z}^{1\prime}\boldsymbol{e}^1 = \mathbf{0}_K$.

Therefore, using the residual errors for (16) and (17):

$$\mathbf{1}_0'\boldsymbol{y}^0 = N(0)\alpha_0^* + \mathbf{1}_0'\mathbf{Z}^0\boldsymbol{\gamma}^* \tag{20}$$

$$\mathbf{1}_1'\boldsymbol{y}^1 = N(1)\alpha_1^* + \mathbf{1}_1'\mathbf{Z}^1\boldsymbol{\gamma}^* \tag{21}$$

If we work out α_0^* and α_1^* from this, we can obtain:

$$\alpha_0^* = \frac{\mathbf{1}_0'\boldsymbol{y}^0}{N(0)} - \frac{\mathbf{1}_0'\mathbf{Z}^0\boldsymbol{\gamma}^*}{N(0)} = \frac{\mathbf{1}_0'\left(\boldsymbol{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*\right)}{N(0)} \tag{22}$$

$$\alpha_1^* = \frac{\mathbf{1}_1'\boldsymbol{y}^1}{N(1)} - \frac{\mathbf{1}_1'\mathbf{Z}^1\boldsymbol{\gamma}^*}{N(1)} = \frac{\mathbf{1}_1'\left(\boldsymbol{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*\right)}{N(1)} \tag{23}$$

(22) and (23) show the quality-constant logarithmic price level. $\mathbf{1}_t'\boldsymbol{y}^t/N(t)$ shows the arithmetic mean of the logarithmic price for time $t = 0, 1$ and $\mathbf{1}_t'\mathbf{Z}^t/N(t)$ shows the arithmetic mean of the characteristic vector for time $t = 0, 1$. In other words, α_0^* is equal to the result obtained by subtracting the average value of all characteristic values from the average value of the logarithmic price (arithmetic average of the quality-adjusted logarithmic price). Based on the above, the hedonic time dummy estimation value based on the logarithmic price change from time 0 to time 1 is the following differential:

$$LP_{HD} = \alpha_1^* - \alpha_0^* \tag{24}$$

The explanatory variable matrices for (18) are expressed as follows:

$$\mathbf{V} = \begin{bmatrix} \mathbf{1}_0 & \mathbf{0}_0 \\ \mathbf{0}_1 & \mathbf{1}_1 \end{bmatrix}, \ \mathbf{Z} = \begin{bmatrix} \mathbf{Z}^0 \\ \mathbf{Z}^1 \end{bmatrix}$$

Here, $\mathbf{V}$ is a $(N(0)+N(1)) \times 2$ matrix and $\mathbf{Z}$ is a $(N(0)+N(1)) \times K$ matrix. If the explanatory variable is rewritten as $\mathbf{X} = \begin{bmatrix} \mathbf{V} & \mathbf{Z} \end{bmatrix}$, since the residual error vector is $\boldsymbol{e} = \boldsymbol{y} - \mathbf{V}\boldsymbol{\alpha}^* - \mathbf{Z}\boldsymbol{\gamma}^*$, the least squares estimator

$$\boldsymbol{\alpha}^* = \begin{bmatrix} \alpha_0^* \\ \alpha_1^* \end{bmatrix} = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\left(\boldsymbol{y} - \mathbf{Z}\boldsymbol{\gamma}^*\right) \tag{25}$$

can be obtained from $\partial \boldsymbol{e}'\boldsymbol{e}/\partial\boldsymbol{\alpha}^* = \mathbf{0}$. Based on (25), the residual error may be rewritten as $\boldsymbol{e} = \mathbf{M}\left(\boldsymbol{y} - \mathbf{Z}\boldsymbol{\gamma}^*\right)$. Here,

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}^0 \\ \mathbf{M}^1 \end{bmatrix} = \mathbf{I} - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}' = \begin{bmatrix} \mathbf{I}_0 - \mathbf{1}_0\mathbf{1}_0'/N(0) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_1 - \mathbf{1}_1\mathbf{1}_1'/N(1) \end{bmatrix},$$

$\mathbf{I}$ is a $(N(0)+N(1)) \times (N(0)+N(1))$ identity matrix, and $\mathbf{I}_t$ is a $N(t) \times N(t)$ identity matrix. If we define $\boldsymbol{y}^* = \mathbf{M}\boldsymbol{y}$ and $\mathbf{Z}^* = \mathbf{M}\mathbf{Z}$, the error sum of squares is $\boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{y}^* - \mathbf{Z}^*\boldsymbol{\gamma}^*\right)'\left(\boldsymbol{y}^* - \mathbf{Z}^*\boldsymbol{\gamma}^*\right)$, so the least squares estimator for $\boldsymbol{\gamma}$ can be obtained as follows:

$$\boldsymbol{\gamma}^* = \left(\mathbf{Z}^{*\prime}\mathbf{Z}^*\right)^{-1}\mathbf{Z}^{*\prime}\boldsymbol{y}^* = \left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)^{-1}\left(\mathbf{Z}^{0*\prime}\boldsymbol{y}^{0*} + \mathbf{Z}^{1*\prime}\boldsymbol{y}^{1*}\right) \tag{26}$$

If we first calculate $\boldsymbol{\gamma}^*$ from (26) and then plug it into (25) (or (22) or (23)), the time effect estimator $\boldsymbol{\alpha}^*$ can be obtained.

2.4.2 Imputed Hedonic Indexes

Instead of performing estimation one time for two periods by pooling data, it is also possible to estimate the characteristic price parameter for each period. Taking $\boldsymbol{\eta}^t$ as the $N(t) \times 1$ error term vector, the regression model for time $t = 0$ and time $t = 1$ may be written as follows:

$$\boldsymbol{y}^t = \mathbf{1}_t\beta_t + \mathbf{Z}^t\boldsymbol{\gamma}^t + \boldsymbol{\eta}^t \tag{27}$$

Here, it is assumed that the characteristic price parameters $\boldsymbol{\gamma}^0, \boldsymbol{\gamma}^1$ vary depending on the observation period. If one includes β_t^* and $\boldsymbol{\gamma}^{t*}$ as least squares estimators, the following formula may be established using the least squares residual error vector $\boldsymbol{u}^t$:

$$\boldsymbol{y}^t = \mathbf{1}_t \beta_t^* + \mathbf{Z}^t \boldsymbol{\gamma}^{t*} + \boldsymbol{u}^t \tag{28}$$

Based on the nature of the residual error, $\begin{bmatrix}\mathbf{1}_t & \mathbf{Z}^t\end{bmatrix}' \boldsymbol{u}^t = \begin{bmatrix}0 & \mathbf{0}_K'\end{bmatrix}'$, so the following can be obtained:

$$\mathbf{1}_0' \boldsymbol{y}^0 = N(0)\beta_0^* + \mathbf{1}_0' \mathbf{Z}^0 \boldsymbol{\gamma}^{0*} \tag{29}$$

$$\mathbf{1}_1' \boldsymbol{y}^1 = N(1)\beta_1^* + \mathbf{1}_1' \mathbf{Z}^1 \boldsymbol{\gamma}^{1*} \tag{30}$$

Therefore, the estimator for (29) and the time effect based on (29) can be obtained by solving these for β_0^*, β_1^*.

$$\beta_0^* = \frac{\mathbf{1}_0' \boldsymbol{y}^0}{N(0)} - \frac{\mathbf{1}_0' \mathbf{Z}^0 \boldsymbol{\gamma}^{0*}}{N(0)} = \frac{\mathbf{1}_0' \left(\boldsymbol{y}^0 - \mathbf{Z}^0 \boldsymbol{\gamma}^{0*}\right)}{N(0)} \tag{31}$$

$$\beta_1^* = \frac{\mathbf{1}_1' \boldsymbol{y}^1}{N(1)} - \frac{\mathbf{1}_1' \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}}{N(1)} = \frac{\mathbf{1}_1' \left(\boldsymbol{y}^1 - \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}\right)}{N(1)} \tag{32}$$

The estimation value of the hedonic time dummy based on the logarithmic price change from time 0 to time 1 may be obtained using the differential $LP_{HD} = \alpha_1^* - \alpha_0^*$. However, since it is assumed that the parameters $\boldsymbol{\gamma}^0, \boldsymbol{\gamma}^1$ for quality adjustment based on the formulation of (27) vary across periods for the two times, it is not possible to simply define the logarithmic price change from the differential of β_0^*, β_1^*. Therefore, Laspeyres type and Paasche type measures of price change in imputed hedonic model are shown as follows :

$$\text{[Laspeyres]} \quad \phi_L^* = \left(\beta_1^* + \frac{\mathbf{1}_1' \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}}{N(1)}\right) - \left(\beta_0^* + \frac{\mathbf{1}_1' \mathbf{Z}^1 \boldsymbol{\gamma}^{0*}}{N(1)}\right) \tag{33}$$

$$\text{[Paasche]} \quad \phi_P^* = \left(\beta_1^* + \frac{\mathbf{1}_0' \mathbf{Z}^0 \boldsymbol{\gamma}^{1*}}{N(0)}\right) - \left(\beta_0^* + \frac{\mathbf{1}_0' \mathbf{Z}^0 \boldsymbol{\gamma}^{0*}}{N(0)}\right) \tag{34}$$

where characteristic value change of a Laspeyres type measure is calculated as the arithmetical mean of $\mathbf{Z}^1$, and that of a Paasche type measure is calculated as an arithmetical mean of $\mathbf{Z}^0$. For both the differential ϕ_L^* and ϕ_P^*, adjustment with characteristic price is asymmetrical. Therefore, using the median value of the two differentials, the hedonic imputed estimation value based on the logarithmic price change from time 0 to time 1 is written as follows:

$$\begin{aligned} LP_{HI} &= \frac{1}{2}\phi_L^* + \frac{1}{2}\phi_P^* \\ &= \frac{\mathbf{1}_1' \left\{\boldsymbol{y}^1 - \mathbf{Z}^1 \left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right\}}{N(1)} - \frac{\mathbf{1}_0' \left\{\boldsymbol{y}^0 - \mathbf{Z}^0 \left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right\}}{N(0)} \end{aligned} \tag{35}$$

Here, one can see that quality adjustment of price is performed not with $\mathbf{Z}^t \boldsymbol{\gamma}^{t*}$ but with $\mathbf{Z}^t \left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)$. If the sample sizes for the two times are identical and the characteristics and characteristic prices are constant over time, there is no difference between the two techniques.

2.4.3 Differences Between Time Dummy Indexes and Imputed Indexes

In order to look at the differences between LP_{HD} (24) and LP_{HI} (35), the differential of the two may be expressed as follows:

$$LP_{HD} - LP_{HI} = \left(\frac{\mathbf{1}_1'\mathbf{Z}^1}{N(1)} - \frac{\mathbf{1}_0'\mathbf{Z}^0}{N(0)}\right)\left(\frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} - \gamma^*\right) \tag{36}$$

In other words, if the average of the characteristic prices is equivalent for each time and if the pooled hedonic regression model characteristic price is equivalent to the hedonic characteristic price median value estimated for each time, (24) and (35) are fully equivalent.

Based on (31) and (32), the β_0, β_1 least squares estimator regressed on each observation period is:

$$\beta_t^* = \mathbf{1}_t'\left(\boldsymbol{y} - \mathbf{Z}^t\boldsymbol{\gamma}^{t*}\right)/N(t)$$

Using this, the least squares residual error may be written as follows:

$$\boldsymbol{u}^t = \mathbf{M}^t\boldsymbol{y}^t - \mathbf{M}^t\mathbf{Z}^t\boldsymbol{\gamma}^{t*} \tag{37}$$

Here, $\mathbf{M}^t = \mathbf{I} - \mathbf{1}_t\mathbf{1}_t'/N(t)$. If we define $\boldsymbol{y}^{t*} = \mathbf{M}^t\boldsymbol{y}^t$ and $\mathbf{Z}^{t*} = \mathbf{M}^t\mathbf{Z}^t$, the estimated characteristic price vector is as follows:

$$\boldsymbol{\gamma}^{t*} = \left(\mathbf{Z}^{t*\prime}\mathbf{Z}^{t*}\right)^{-1}\mathbf{Z}^{t*\prime}\boldsymbol{y}^{t*} \tag{38}$$

Here, if we multiply $\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)$ by both sides of (26), the characteristic price using pooled data in (26) becomes:

$$\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)\gamma^* = \left(\mathbf{Z}^{0*\prime}\boldsymbol{y}^{0*} + \mathbf{Z}^{1*\prime}\boldsymbol{y}^{1*}\right) = \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{1*} \tag{39}$$

Note that, based on (38), $(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*})\gamma^{0*} = \mathbf{Z}^{0*\prime}\boldsymbol{y}^{0*}$ and $(\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*})\gamma^{1*} = \mathbf{Z}^{1*\prime}\boldsymbol{y}^{1*}$. If $\boldsymbol{\gamma}^{0*}$ and $\boldsymbol{\gamma}^{1*}$ are equivalent, (39) shows that $\boldsymbol{\gamma}^*$ is necessarily the shared characteristic vector of these. If we multiply the right side of (36) $\left(\frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} - \gamma^*\right)$ by $2\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)$ from the right side, we obtain the following:

$$\begin{aligned}
&2\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)\left(\frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} - \gamma^*\right)\\
&= \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{1*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{1*} - 2\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)\gamma^*\\
&= \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{1*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{1*} - 2\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{0*} - 2\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{1*}\\
&= -\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\gamma^{1*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{0*} - \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\gamma^{1*}\\
&= -\left(\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*} - \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\right)\left(\gamma^{1*} - \gamma^{0*}\right)
\end{aligned}$$

In other words,

$$\left(\frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} - \gamma^*\right) = -\frac{1}{2}\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)^{-1} \times \left(\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*} - \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\right)\left(\gamma^{1*} - \gamma^{0*}\right) \tag{40}$$

If we plug (40) into (36), the differential of the time dummy index and imputed index shown with the logarithmic price difference may be rewritten using the following formula:

$$\begin{aligned}
&LP_{HD} - LP_{HI}\\
&= -\frac{1}{2}\left(\frac{\mathbf{1}_1'\mathbf{Z}^1}{N(1)} - \frac{\mathbf{1}_0'\mathbf{Z}^0}{N(0)}\right)\left(\mathbf{Z}^{0*\prime}\mathbf{Z}^{0*} + \mathbf{Z}^{1*\prime}\mathbf{Z}^{1*}\right)^{-1} \times \left(\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*} - \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}\right)\left(\gamma^{1*} - \gamma^{0*}\right)
\end{aligned} \tag{41}$$

Based on the above, when any of the following conditions are met, the two logarithmic price differentials based on the hedonic time dummy and hedonic imputation method are identical:

- The average value of each characteristic is equivalent for the two times: $\frac{\mathbf{1}_1'\mathbf{Z}^1}{N(1)} = \frac{\mathbf{1}_0'\mathbf{Z}^0}{N(0)}$
- The characteristic value variance/covariance matrices are equivalent for the two times: $\mathbf{Z}^{1*\prime}\mathbf{Z}^{1*} = \mathbf{Z}^{0*\prime}\mathbf{Z}^{0*}$
- The quality-adjusted prices obtained with the hedonic price estimation method for each time are identical: $\boldsymbol{\gamma}^{1*} = \boldsymbol{\gamma}^{0*}$

2.4.4 Summary of Hedonic Dummy Indexes and Hedonic Imputed Indexes

As shown above, we have identified factors that show the differences between a hedonic dummy index and hedonic imputed index. In the regression equations, if it is possible to use information for two times and formulate the indexes with identical function forms, taking the (geometric) average of the two is perhaps a viable method when the two approaches show different results. However, rather than doing this, using either one index or the other is preferable for various reasons.

A major issue of concern when using the hedonic time dummy (HD) method is that it has the following restriction: the characteristic price is fixed over time. However, the null hypothesis that the characteristic variable parameter is fixed throughout the observation period has in fact been dismissed by a number of papers. In contrast to this, the hedonic imputed index (HI) method is inherently more flexible than the time dummy model, which is a significant advantage.

In Section 2.4.3, we showed that the difference between the two approaches depends on the following three variable factors:

- The characteristic average value
- The variance/covariance matrix of the characteristic value
- The estimated hedonic characteristic price

What's more, multiplication of the difference between the two periods produces the ultimate difference. Therefore, the stability of the characteristic price parameter alone is not necessarily a problem. For example, even if the parameter is unstable, its instability will be alleviated by slight changes in other characteristics, and the same may be true for the price index.

Due to the nature of the HD method, it uses independent variables observed for the two times, and it is restricted such that the characteristic price parameters are the same for the two times, and regression analysis ends up being executed one time only. In this sense, it may be said that the HD method is not flexible due to the presence of these restrictions. Why, then, are these restrictions imposed? Presumably, the reasons include the following:

- To not lose a degree of freedom.
- To provide an unambiguous estimation value for the overall price change from time 0 to time 1.
- To minimize the effect of abnormal values in conditions where there is a small degree of freedom.

In contrast to this, the HI method allows for diachronic changes in characteristic prices and formulation is more flexible. However:

- A degree of freedom is lost.
- The estimation value for the overall price change in the two times is difficult to repro-

duce.

Due to these and other issues, analysis costs increase. The latter of the two issues pointed out above may in fact not be all that serious, because Laspeyres- and Paasche-type estimation values for price changes are well established in index theory. Bearing these points in mind, the HI method may be considered the preferred method as long as the degree of freedom is not extremely restricted.

In light of the above, "*the rolling window hedonic method*" that merges the hedonic dummy method and hedonic imputed method has been proposed. Market structural changes occur as a result of various exogenous shocks, but it is thought that a certain adjustment period exists until shocks are absorbed by the market and changes are realized. Therefore, the regression coefficient likewise does not change instantaneously but should instead be viewed as changing sequentially.

However, if estimating a model where the data is divided into various periods and observation data for each period is used (as with the HI method), the links to prior and subsequent data are severed. As a result, under the assumption that structural changes occur sequentially, this method ends up making it more difficult to capture price changes within the sequential change process. Instead, as a more natural approach, a method of estimating price indexes within the sequential change process by taking an estimation period of a certain duration and estimating the model while moving this period – as if obtaining a moving average – may be preferable. A method that has been proposed based on this idea is "*the rolling window hedonic method*". This approach is employed in the estimation of housing price indexes in Ireland and Japan.

2.5 Hedonic Production Price Index Measurement and Quality Adjustment

2.5.1 The Producer Revenue Maximization Problem

In this section, we will explain the characteristics of quality-adjusted hedonic indexes for producer prices, based on Diewert (2002)[24]. The Konus-type price index proposed in that paper is defined using a revenue function that is a value function of the revenue maximization problem based on company technology and resource constraints. The revenue function is derived from characteristic values constituting the product price, production technology, production factors, and product.

We shall define the hedonic price (producer's willingness to pay) based on the characteristic vector $\boldsymbol{z}$ as:

$$\Pi^t(\boldsymbol{z}) = \rho^t f^t(\boldsymbol{z}) \tag{42}$$

Here, ρ^t is the price showing the value of all characteristic values used for the product at time t and $f^t(\boldsymbol{z})$ shows the cardinal utility separable from the utility function. In (42), it is assumed that the utility function is equivalent for the two times.

$$f^0 = f^1 \tag{43}$$

Given the hedonic price (42), the producer performs revenue maximization. First, we shall define the production function F as follows:

$$q = F^t(\boldsymbol{z}, \boldsymbol{v}) \tag{44}$$

Here, q is the production volume and $\boldsymbol{v}$ is the production factor vector. For a given level of production technology, the following revenue-maximizing value function may be obtained:

$$\begin{aligned} R(\rho^s f^s, F^t, \boldsymbol{Z}^t, \boldsymbol{v}) &\equiv \max_{q,\boldsymbol{z}} \left\{ \rho^s f^s(\boldsymbol{z}) q : q = F^t(\boldsymbol{z}, \boldsymbol{v}); \boldsymbol{z} \in \boldsymbol{Z}^t \right\} \\ &= \max_{\boldsymbol{z}} \left\{ \rho^s f^s(\boldsymbol{z}) F^t(\boldsymbol{z}, \boldsymbol{v}); \boldsymbol{z} \in \boldsymbol{Z}^t \right\} \end{aligned} \tag{45}$$

Here, $\boldsymbol{Z}^t$ shows the feasible set of characteristic values.

When the characteristics and input factors for time t are taken as $\boldsymbol{z}^t, \boldsymbol{v}^t$, the corresponding production volume is:

$$q^t = F^t(\boldsymbol{z}^t, \boldsymbol{v}^t) \tag{46}$$

Therefore, the maximized revenue function for time t may be written as follows:

$$\begin{aligned} R(\rho^t f^t, F^t, \boldsymbol{Z}^t, \boldsymbol{v}^t) &\equiv \max_{q,\boldsymbol{z}} \left\{ \rho^t f^t(\boldsymbol{z}) q : q = F^t(\boldsymbol{z}, \boldsymbol{v}^t); \boldsymbol{z} \in \boldsymbol{Z}^t \right\} \\ &= \rho^t f^t(\boldsymbol{z}^t) q^t; \quad t = 0, 1 \end{aligned} \tag{47}$$

2.5.2 Konus-Type Hedonic Production Price Indexes

Using the maximized revenue function (47), the Konus-type hedonic product price index between time 0 and time 1 is defined as follows:

$$P(\rho^0 f^0, \rho^1 f^1, F^t, \boldsymbol{Z}^t, \boldsymbol{v}) = \frac{R(\rho^1 f^1, F^t, \boldsymbol{Z}^t, \boldsymbol{v})}{R(\rho^0 f^0, F^t, \boldsymbol{Z}^t, \boldsymbol{v})} \tag{48}$$

The differences between the two revenue functions are caused by the hedonic prices $\rho^1 f^1$ and $\rho^0 f^0$. Since $\max_{\boldsymbol{z}} \left\{ \rho^1 f^1(\boldsymbol{z}) F^t(\boldsymbol{z}, \boldsymbol{v}^t); \boldsymbol{z} \in \boldsymbol{Z}^t \right\} = \max_{\boldsymbol{z}} \left\{ \rho^1 f^0(\boldsymbol{z}) F^t(\boldsymbol{z}, \boldsymbol{v}^t); \boldsymbol{z} \in \boldsymbol{Z}^t \right\}$ based on hypothesis (43), (48) may be rewritten as follows:

$$P(\rho^0 f^0, \rho^1 f^1, F^t, \boldsymbol{Z}^t, \boldsymbol{v}) = \frac{\rho^1 R(\rho^0 f^0, F^t, \boldsymbol{Z}^t, \boldsymbol{v})}{\rho^0 R(\rho^0 f^0, F^t, \boldsymbol{Z}^t, \boldsymbol{v})} = \frac{\rho^1}{\rho^0} \tag{49}$$

In estimation of the hedonic price, if we assume that the utility of the characteristic portion is diachronically constant, the Konus-type product price index may be estimated very easily.

Let us consider general cases that do not meet hypothesis (43). Taking the price index in (49) as our base, we can define an observable hedonic Laspeyres production price index and Paasche production price index with the following inequalities, using:

$$P(\rho^0 f^0, \rho^1 f^1, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0) = \frac{R(\rho^1 f^1, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0)}{R(\rho^0 f^0, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0)} \geq \frac{\rho^1 f^1(\boldsymbol{z}^0)}{\rho^0 f^0(\boldsymbol{z}^0)} = P_{HL} \tag{50}$$

$$P(\rho^0 f^0, \rho^1 f^1, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1) = \frac{R(\rho^1 f^1, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1)}{R(\rho^0 f^0, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1)} \leq \frac{\rho^1 f^1(\boldsymbol{z}^1)}{\rho^0 f^0(\boldsymbol{z}^1)} = P_{HP} \tag{51}$$

Here, $P(\rho^0 f^0, \rho^1 f^1, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0)$ and $P(\rho^0 f^0, \rho^1 f^1, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1)$ are theoretical production price indexes that cannot be observed. (50) shows that the theoretical production price index $P(\rho^0 f^0, \rho^1 f^1, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0)$ has the observable Laspeyres production price index P_{HL} as its lower limit, and (51) shows that the theoretical production price index $P(\rho^0 f^0, \rho^1 f^1, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1)$ has the observable Paasche production price index P_{HP} as its upper limit.

By using these convex combination equations (weighted averages) instead of $F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0$ or $F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1$ constituting the production price index, it is possible to define the range that can be covered by the theoretical Laspeyres production price index and Paasche production price index. If the scalar $\lambda \in [0,1]$ is used, the convex combinations for $F^t, \boldsymbol{Z}^t, \boldsymbol{v}^t$ for period $t = 0, 1$ may be written as follows.

$$\begin{aligned}
\boldsymbol{Z}(\lambda) &= (1-\lambda)\boldsymbol{Z}^0 + \lambda \boldsymbol{Z}^1 \\
\boldsymbol{v}(\lambda) &= (1-\lambda)\boldsymbol{v}^0 + \lambda \boldsymbol{v}^1 \\
F(\lambda) &= (1-\lambda)F^0(\boldsymbol{z}, \boldsymbol{v}(\lambda)) + \lambda F^1(\boldsymbol{z}, \boldsymbol{v}(\lambda))
\end{aligned}$$

Therefore, the hedonic production price function may be written as:

$$P(\lambda) = \frac{R(\rho^1 f^1, F(\lambda), \boldsymbol{Z}(\lambda), \boldsymbol{v}(\lambda))}{R(\rho^0 f^0, F(\lambda), \boldsymbol{Z}(\lambda), \boldsymbol{v}(\lambda))} = \frac{\max_{\boldsymbol{z}} \left\{\rho^1 f^1(\boldsymbol{z}) F(\lambda); \boldsymbol{z} \in \boldsymbol{Z}(\lambda)\right\}}{\max_{\boldsymbol{z}} \left\{\rho^0 f^0(\boldsymbol{z}) F(\lambda); \boldsymbol{z} \in \boldsymbol{Z}(\lambda)\right\}} \tag{52}$$

When $\lambda = 0,$ since $P(\lambda)$ signifies that $P(\rho^0 f^0, \rho^1 f^1, F^0, \boldsymbol{Z}^0, \boldsymbol{v}^0),$ the following may be derived from inequality (50):

$$P(0) \geq P_{HL} = \frac{\rho^1 f^1(\boldsymbol{z}^0)}{\rho^0 f^0(\boldsymbol{z}^0)} \tag{53}$$

As well, when $\lambda = 1$, since $P(\lambda)$ signifies that $P(\rho^0 f^0, \rho^1 f^1, F^1, \boldsymbol{Z}^1, \boldsymbol{v}^1)$, the following may be derived from inequality (51):

$$P(1) \leq P_{HP} = \frac{\rho^1 f^1(\boldsymbol{z}^1)}{\rho^0 f^0(\boldsymbol{z}^1)} \tag{54}$$

By using Diewert's proof (1983; 1060-1061)[22], if $P(\lambda)$ is a continuous function for $[0,1]$, it is possible to show that λ^* exists, whereby

$$0 \leq \lambda^* \leq 1 \text{ and } P_{HL} \leq P(\lambda^*) \leq P_{HP} \quad P_{HP} \leq P(\lambda^*) \leq P_{HL}.$$

In other words, one can see that the theoretical hedonic production price index for the period $t = 0, 1$, when considered via $P(\lambda^*)$ described above, exists between the observable Laspeyres production price index and Paasche production price index. Note that to obtain this result, one must assume the continuity of λ in the hedonic model price functions $\rho^1 f^1(\boldsymbol{z}^0), \rho^0 f^0(\boldsymbol{z}^0)$ in the numerator and denominator of Formula (52), the production functions $F^0(\boldsymbol{z}, \boldsymbol{v}), F^1(\boldsymbol{z}, \boldsymbol{v})$, and the feasible characteristic value sets $\boldsymbol{Z}^0, \boldsymbol{Z}^1$. The sufficient conditions for continuity are:

- The production functions $F^0(\boldsymbol{z}, \boldsymbol{v}), F^1(\boldsymbol{z}, \boldsymbol{v})$ are positive and continuous for $\boldsymbol{z}$ and $\boldsymbol{v}$.
- The hedonic model price functions $f^0(\boldsymbol{z}), f^1(\boldsymbol{z})$ are positive and continuous for $\boldsymbol{z}$.
- ρ^0, ρ^1 are positive.
- Sets $\boldsymbol{Z}^0, \boldsymbol{Z}^1$ are convex sets, bounded, and closed.

Based on the above, one can see that the boundary range for the theoretical price index is determined by the observable price index. In order to obtain the best value for approximating the theoretical index, it is natural to take the adjusted average of the two boundary values. If the adjusted average function for the Laspeyres and Paasche production price indexes is written as $m(P_{HL}, P_{HP})$, we can confirm, based on Diewert's argument (1997; 138)[23], that $m()$ must be the geometric average. In other words, the best candidate in terms of approximating the theoretical production price index is the following observable Fisher hedonic production price index, using (50) and (51):

$$P_{HF} = (P_{HL} P_{HP})^{1/2} = \frac{\rho^1}{\rho^0} \left(\frac{f^1(\boldsymbol{z}^0)}{f^0(\boldsymbol{z}^0)}\right)^{1/2} \left(\frac{f^1(\boldsymbol{z}^1)}{f^0(\boldsymbol{z}^1)}\right)^{1/2}$$

If the hypothesis $f^0 = f^1$ is fulfilled by the hedonic model price function being the same for the two times, then this can be transformed into $P_{HF} = \rho^1/\rho^0$. As well, if the respective observable prices are defined as

$$P^0 = \rho^0 f^0(\boldsymbol{z}^0) \text{ and } P^1 = \rho^1 f^1(\boldsymbol{z}^1) \tag{55}$$

the Laspeyres and Paasche production price indexes can be shown as quality-adjusted price comparisons:

$$P_{HL} = \frac{\rho^1 f^1(\boldsymbol{z}^0)}{\rho^0 f^0(\boldsymbol{z}^0)} = \frac{P^1/f^1(\boldsymbol{z}^1)}{P^0/f^1(\boldsymbol{z}^0)} \tag{56}$$

$$P_{HP} = \frac{\rho^1 f^1(\boldsymbol{z}^1)}{\rho^0 f^0(\boldsymbol{z}^1)} = \frac{P^1/f^0(\boldsymbol{z}^1)}{P^0/f^0(\boldsymbol{z}^0)} \tag{57}$$

Therefore, the Fisher hedonic production price index may be written as follows:

$$P_{HF} = (P_{HL}P_{HP})^{1/2} = \left(\frac{P^1/f^1(\boldsymbol{z}^1)}{P^0/f^1(\boldsymbol{z}^0)}\right)^{1/2} \left(\frac{P^1/f^0(\boldsymbol{z}^1)}{P^0/f^0(\boldsymbol{z}^0)}\right)^{1/2} \tag{58}$$

In other words, the Fisher hedonic production price index may be obtained from the geometric average of the two quality-adjusted price indexes obtained by estimating the hedonic regression model. The hedonic approach is useful not just for quality adjustment of the product user price but also for quality adjustment of the product supplier price. In this chapter, in order to define a product price index assuming competitive company production activities, we used a revenue function (total willingness to pay) maximized based on Konus. If the cardinal utility function for the characteristic portion is the same at the two points in time, the theoretical production price index may be shown by comparison with the observable product price.

In addition, in general cases, based on certain restrictions, we showed that the theoretical production price index is present in the range that forms the boundary values of the observable Laspeyres and Paasche production price indexes.

2.6 Characteristics, Advantages, and Disadvantages of the Hedonic Method

Rosen (1974)[55] developed a market equilibrium theory for differentiated products. This study rigorously examined the relationship between the structures of the product supplier offer function, product demander bid value function, and hedonic market price function, and characterized product market price based on consumer and producer behavior.

If the bid value function is used, it is possible to obtain the consumer's willingness to pay with respect to changes in characteristics. In market equilibrium, not only are the market price and bid value consistent, but the slope of the hedonic function and bid value function are also consistent. Since the bid value function is a concave function, the willingness to pay with respect to incremental changes in characteristics is smaller than the change in the market price. In other words, caution is required with respect to the market price function characteristic limit value, since it is possible that the willingness to pay will be overestimated as long as demanders are not homogeneous. However, if one assumes that changes in characteristic values will be sufficiently small, the market price function limit value may be used as an approximation of the willingness to pay. Therefore, the market price function is generally estimated in existing research.

If the hedonic approach is used, it is possible to measure changes in quality-adjusted price using samples at another point in time. The simplest and most widely used method is to estimate the time effect for the hedonic function based on the characteristic price being constant and using a time dummy with pooled data. Hedonic imputed indexes estimate a hedonic function for each observation time, allowing for changes in characteristic prices, and measure price changes using a Laspeyres-type scale or Paasche-type scale. Hedonic dummy indexes and hedonic imputed indexes produce different results, but this is not caused solely by differences based on whether or not characteristic prices are constant for the two times. Differences between the two indexes occur if the average value for each characteristic varies for the two times or if the characteristic variance /covariance matrices vary for the two periods.

The hedonic approach is useful not just for quality adjustment of the product user price but also for quality adjustment of the product supplier price. It is possible to define a producer price index using a revenue function (total willingness to pay) maximized based on Konus. In this case, if the cardinal utility function for the characteristic portion is the same at the two points in time, the theoretical production price index may be shown by comparison with the observable product price. In general cases, based on certain restrictions, it has been shown that the theoretical production price index is present in the range that forms the boundary values of the observable Laspeyres and Paasche production price indexes.

Thus, one can see that price indexes with what is broadly called the "hedonic method" vary considerably depending on the approach employed in estimation.

The advantages and disadvantages of the hedonic method in the estimation of property price indexes are outlined below. The following may be considered as advantages:

- As well as having a basis in economic theory and index theory, the theoretical biases of the hedonic method are clear.
- Compared to other approaches, it is possible to use all transaction price data, so it may be considered the most efficient approach.
- Since it makes it possible to control for the many characteristics of property, it enables the sorting of data into specialized indexes by purpose/region.
- Since it is already used in the estimation of consumer price statistics and the like, it is possible to be consistent with other economic statistics.

Disadvantages include:

- Since it is necessary to collect many property-related characteristics, information-gathering costs are high.
- In cases where it is not possible to collect important characteristics for determining property prices, one faces the problem of omitted variable bias.
- Calculated indexes vary depending on the function form used. In other words, there is a low level of reproducibility.
- In cases of strong heterogeneity, it may not be possible to control for quality.
- Since the underlying economic theory and statistical procedures are complicated, the organizations creating the indexes require specialized skills, and explaining the indexes to users is difficult.

3 Repeat Sales Method

3.1 Standard Repeat Sales Price Index

3.1.1 Regression for Repeat Sales

Apart from the hedonic method, the most used approach is the repeat sales method elaborated by Bailey, Muth, and Nourse (1963)[2] and Case and Shiller (1987)[9], (1989)[10]. With the repeat sales method, since use of the data generation process in the hedonic price regression model is assumed, some of the problems occurring with the hedonic method are inherited. However, since the same product is compared, underestimation bias is eliminated if there is no change in the characteristics or characteristic prices. Given that the estimation method is straightforward, it has the benefits of being a technique with high reproducibility and estimation efficiency.

With either method, there is a bias that exists due to the estimation technique. Since the purpose of a price index is to observe price data over an extended time, as the observation period becomes longer, "aggregation bias" is to be expected, due to changes in the characteristics and characteristic prices of identical properties.

In particular, the fact that it is not possible to separate effects common to the market as a whole (time effects) that are factors in the housing market supply-demand balance from effects related to changes in individual housing, especially deterioration (age effects), is an extremely important issue when using the repeat sales method. If the effects of housing deterioration are ignored, it is to be expected that repeat sales price indexes will have a strong downward bias.

As well, since only properties transacted multiple times are selected for use, the sample size shrinks and there is also concern that selection bias occurs in the samples. Moreover, while it is strongly assumed that there will be no changes in property quality during the period when repeat transactions are conducted, it is easy to predict that property deterioration, investment in renovations, or changes to the surrounding environment will occur, so the assumption is not consistent with the reality.

Below, we provide an overview of the repeat sales method and describe what kinds of problems occur with it. Section 3.2 explains the analysis structure and price index characteristics of the standard repeat sales method. Section 3.3 and Section 3.4 present the problems of aggregation bias and sample selection bias with the standard repeat sales method, along with methods of resolving them.

When housing prices at multiple points in time are observed as data, the hedonic market price of property n at time t may be described in the form of the following regression model:

$$y_n^t \equiv \ln p_n^t = \alpha_t + \boldsymbol{z}_n^{t\prime}\boldsymbol{\gamma}^t + \varepsilon_n^t \quad (n = 1, 2, \ldots, N(t);\ t = 0, 1, \ldots, T) \tag{59}$$

Here, y_n^t is the housing price logarithm ($\ln p_n^t$), α_t is the unknown time effect at time t, $\boldsymbol{z}_n^t = (1, z_{n2}^t, \ldots, z_{nk}^t, \ldots, z_{nK}^t)'$ is the explanatory variable (characteristic) vector including a constant term, $\boldsymbol{\gamma}^t = (\gamma_1, \gamma_2^t, \ldots, \gamma_k^t, \ldots, \gamma_K^t)'$ is the unknown coefficient vector, and ε_n^t is the error term. γ_1 is the constant term coefficient for the overall model and $\gamma_2^t, \ldots, \gamma_K^t$ is the characteristic marginal effect (characteristic quality-adjusted parameter).

Let us take it that property n is transacted on the market twice, at time s and time t ($t > s$). In this case, for example, the logarithmic price differential for the two times in (59) may be written as follows:

$$Y_n \equiv \ln p_n^t - \ln p_n^s = (\alpha_t - \alpha_s) + \left(\boldsymbol{z}_n^{t\prime}\boldsymbol{\gamma}^t - \boldsymbol{z}_n^{s\prime}\boldsymbol{\gamma}^s\right) + v_n \tag{60}$$

Here, v_n is the differential of the error terms for the respective times ($\varepsilon_n^t - \varepsilon_n^s$). In other words, the price rate of change (logarithmic differential) may be treated as data occurring based on differences in the time effect, changes in characteristic values (characteristic qualities and quantities), and errors.

The repeat sales method of Bailey, Muth, and Nourse (1963)[2] and Case and Shiller (1987)[9], (1989)[10] reformulated (60) above by implicitly establishing the following hypotheses:

Hypothesis 1. All characteristics are constant over time.

Hypothesis 2. All characteristic parameters are constant over time.

In other words, hypothesis 1 signifies that $\boldsymbol{z}_n = \boldsymbol{z}_n^t = \boldsymbol{z}_n^s$ and hypothesis 2 that $\boldsymbol{\gamma} = \boldsymbol{\gamma}^t = \boldsymbol{\gamma}^s$. If one takes property n as being transacted for the first time at time s and for the second time at time t, the hedonic regression formula (59) may be rewritten as follows for time s and time t, respectively, using time dummy variables based on hypothesis 1 and 2:

$$y_n^s = \overline{\boldsymbol{d}}_n' \boldsymbol{\alpha} + \boldsymbol{z}_n' \boldsymbol{\gamma} + \varepsilon_n^s \tag{61}$$

$$y_n^t = \overline{\overline{\boldsymbol{d}}}_n' \boldsymbol{\alpha} + \boldsymbol{z}_n' \boldsymbol{\gamma} + \varepsilon_n^t \tag{62}$$

Note that $\overline{\boldsymbol{d}}_n' = (\overline{d}_n^1, \ldots, \overline{d}_n^T)$ is the time dummy variable for the first transaction and $\overline{\overline{\boldsymbol{d}}}_n' = (\overline{\overline{d}}_n^1, \ldots, \overline{\overline{d}}_n^T)$ is the time dummy variable for the second transaction, and they are defined as follows:

$$\overline{d}_n^u = \begin{cases} 1 & u = s \\ 0 & u \neq s \end{cases}, \quad \overline{\overline{d}}_n^u = \begin{cases} 1 & u = t \\ 0 & u \neq t \end{cases}$$

As well, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s, \ldots, \alpha_t, \ldots, \alpha_T)$ is the time effect vector. Since there is a linear relationship between the constant term $\boldsymbol{z}^1$ for the overall model and the dummy variable, the time effect for time 0 is standardized here as $\alpha_0 = 0$. Therefore, the time dummy variables $\overline{d}_n^0$ and $\overline{\overline{d}}_n^0$ for time 0 are omitted.

The differential of the time dummy variables for the first and second hedonic regression equations is defined with the following $T \times 1$ vector.

$$\boldsymbol{D}_n = \overline{\overline{\boldsymbol{d}}}_n - \overline{\boldsymbol{d}}_n \tag{63}$$

Note that:

$$D_n^u = \begin{cases} 1 & u = t \text{ (2nd transaction)} \\ -1 & u = s \text{ (1st transaction)} \\ 0 & \text{other cases} \end{cases} \qquad (n = 1, 2, \ldots, N(t);\ u = 1, \ldots, s, \ldots, t, \ldots, T)$$

$\boldsymbol{Y} = (Y_1, \ldots, Y_n, \ldots, Y_N)'$ and $\mathbf{D} = (\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n, \ldots, \boldsymbol{D}_N)'$ are included, the matrix representation repeat sales regression model may be defined as follows:

$$\boldsymbol{Y} = \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{v} \tag{64}$$

The least squares estimator for (64) is $\hat{\boldsymbol{\alpha}} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\boldsymbol{Y}$. The theoretical value (logarithmic price differential) of a random property transacted for the first time at time s and the second time at time t is:

$$\hat{\boldsymbol{Y}} = \ln\left(\frac{p_n^t}{p_n^s}\right) = \hat{\alpha}_t - \hat{\alpha}_s$$

Therefore, taking time s as the baseline, the price index at time t (price comparison) is $\exp(\hat{\alpha}_t - \hat{\alpha}_s)$. Since the time dummy variable for time 0 was omitted and $\alpha_0 = 0$ included in order to avoid multicollinearity, $\exp(\hat{\alpha}_t)$ is the price index taking time 0 as the baseline. The "BMN" price index presented in Bailey, Muth, and Nourse (1963)[2] is:

$$\boldsymbol{I}^{BMN} = \{\exp(\hat{\alpha}_t), \exp(\hat{\alpha}_1), \ldots, \exp(\hat{\alpha}_T)\} \tag{65}$$

3.1.2 Random Walk Error Term

The following is assumed with respect to the error terms in (61) and (62).

$$\begin{array}{ll} E(\varepsilon_n^t) = 0,\ E\left[(\varepsilon_n^t)^2\right] = \sigma^2 & n = 1, \ldots, N. \\ E(\varepsilon_m^s \varepsilon_n^t) = 0 & m, n = 1, \ldots, N;\ s, t = 1, \ldots, T;\ n \neq m, t \neq s \end{array} \tag{66}$$

(66) shows that for the respective hedonic regression equations, the error terms are homogeneously variant and there is no serial correlation. In this case, since the error term for (6), which is the differential of (3) and (4), is

$$\begin{array}{ll} E(v_n) = 0,\ E\left[(v_n)^2\right] = 2\sigma^2 & n = 1, \ldots, N. \\ E(v_n v_m) = 0 & m, n = 1, \ldots, N;\ n \neq m \end{array} \tag{67}$$

(67) fulfills the conditions of being homogeneously variant and having no serial correlation. Bailey, Muth, and Nourse (1963)[2]'s price index is also estimated based on this type of assumption.

With regard to this, Case and Shiller (1987)[9], (1989)[10] presented a repeat sales regression model which assumes that as the interval between transactions becomes larger, the variance in noise associated with housing-specific structural factors becomes greater, and logarithmic price changes are not homogeneously variant. In this paper, the error term with respect to logarithmic price fluctuation is hypothesized with the following formulas including a random walk:

$$\varepsilon_n^t = h_n^t + \nu_n^t,\ \nu_n^t \sim \text{i.i.d.}\, N(0, \sigma_\nu^2) \quad n = 1, \ldots, N;\ t = 1, \ldots, T \tag{68}$$

$$h_n^t = h_n^{t-1} + \eta_n^t,\ \eta_n^t \sim \text{i.i.d.}\, N(0, \sigma_\eta^2) \quad n = 1, \ldots, N;\ t = 1, \ldots, T \tag{69}$$

Here, the left-side first term in (68) is the random walk shown in (69), and the second item in (68) is assumes white noise ν_n^t, η_n^t:

$$\begin{array}{ll} E(\nu_n^t \nu_m^s) = 0, & n \neq m = 1, \ldots, N;\ t \neq s = 1, \ldots, T \\ E(\nu_n^t \eta_m^s) = 0, & n = 1, \ldots, N;\ t = 0, \ldots, T \\ E(\eta_n^t \eta_m^s) = 0, & n \neq m = 1, \ldots, N;\ t \neq s = 0, \ldots, T \end{array} \tag{70}$$

(64) is the repeat sales regression model error term. Here, based on

$$v_n = \varepsilon_n^t - \varepsilon_n^s = (h_n^t - h_n^s) + (\nu_n^t - \nu_n^s),$$

the following can be obtained:

$$E(v_n) = 0 \tag{71}$$

$$E({v_n}^2) = 2\sigma_\nu^2 + (t - s)\sigma_\eta^2 \tag{72}$$

In this case, one can see that if the transaction interval $t - s$ becomes greater, the repeat sales regression model error variance also increases (heterogeneous variance).

With regard to this heterogeneous variance, Case and Shiller (1987)[9], (1989)[10] proposed a three-step estimation method, the Weighted Repeat Sales (WRS) method.

1. (64) is estimated in the same way as when a BMN price index is obtained, the logarithmic price differential is regressed on the time dummy differential, and the least squares residual error $\widehat{v}_n$ is obtained.
2. In order to estimate $\sigma_\nu^2, \sigma_\eta^2$ in (72), the error value of squares $\widehat{v}_n^2$ is regressed on the constant term and transaction interval $A_n = t - s$. ($\widehat{v}_n^2 = a + bA_n +$ error$_n$)
3. Taking the theoretical value in Step 2 as $\widehat{\widehat{v}}_n^2 = \widehat{a} + \widehat{b}A_n$ and its reciprocal square root $1/\widehat{\widehat{v}}_n^2$ as the weight, the weighted method of least squares is implemented for (64).

If the weight in Step 3 ($N \times N$ diagonal matrix) is defined as

$$\widehat{\omega} = \begin{pmatrix} 1/\widehat{\widehat{v}}_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\widehat{\widehat{v}}_N^2 \end{pmatrix},$$

the weighted repeat sales regression model may be written as follows:

$$\boldsymbol{Y}^* = \mathbf{D}^*\boldsymbol{\alpha} + \boldsymbol{v}^* \tag{73}$$

Note that $\boldsymbol{Y}^* = \widehat{\omega}\boldsymbol{Y}$, $\mathbf{D}^* = \widehat{\omega}\mathbf{D}$, and $\boldsymbol{v}^* = \widehat{\omega}\boldsymbol{v}$. Therefore, a workable generalized least squares estimator is:

$$\widehat{\boldsymbol{\alpha}}^{WLS} = (\mathbf{D}'\widehat{\omega}'\widehat{\omega}\mathbf{D})^{-1}\mathbf{D}'\widehat{\omega}'\widehat{\omega}\boldsymbol{Y} \tag{74}$$

Based on (16), Case and Shiller's WRS price index is:

$$\boldsymbol{I}^{WRS} = \{\exp(0), \exp(\widetilde{\alpha}_1), \ldots, \exp(\widetilde{\alpha}_T)\} \tag{75}$$

Hill, Knight, and Sirmans (1997)[59] and , and Knight, Hill and Sirmans (1999)[43] define the hedonic price index error term in cases where the serial correlation autoregressive parameter is 1 as:

$$\varepsilon_n^t = \varepsilon_n^{t-1} + \nu_n^t,\ \nu_n^t \sim \text{i.i.d.}\, N(0, \sigma_{\nu,n}^2) \quad n = 1, \ldots, N;\ t = 0, \ldots, T$$

Here, ν_n^t assumes unknown heterogeneous variance. Therefore, the repeat sales regression model error variance is:

$$E[(\varepsilon_n^t - \varepsilon_n^s)^2] = (t - s)\sigma_{\nu,n}^2$$

When the transaction interval is taken as $A_n = t - s$, then

$$v_n \equiv \varepsilon_n^t - \varepsilon_n^s \sim N(0, A_n\sigma_{\nu,n}^2),$$

and the weighted repeat sales regression model may be written as follows:

$$\boldsymbol{Y}^{**} = \mathbf{D}^{**}\boldsymbol{\alpha} + \boldsymbol{v}^{**} \tag{76}$$

Note that $\boldsymbol{Y}^{**} = \omega\boldsymbol{Y}$, $\mathbf{D}^{**} = \omega\mathbf{D}$, $\boldsymbol{v}^{**} = \omega\boldsymbol{v}$, and

$$\omega = \begin{pmatrix} 1/\sqrt{A_1} & & & & 0 \\ & \ddots & & & \\ & & 1/\sqrt{A_n} & & \\ & & & \ddots & \\ 0 & & & & 1/\sqrt{A_N} \end{pmatrix}$$

3.2 Aggregation Bias

3.2.1 Bias Due to Omission of the Age Effect

With the repeat sales method, aggregation bias is an important issue that has been pointed out in much research. Aggregation bias is a problem relating to the two hypotheses in repeat sales regression model (60) (hypothesis 1: all characteristics are constant over time; hypothesis 2: all characteristic parameters are constant over time). For example, if there are changes in characteristic values due to deterioration/obsolescence of real estate capital, renovations and maintenance, changes in the surrounding geographic environment, etc., such hypotheses are not valid. As well, Dombrow, Knight, and Sirmans (1997)[41] tested whether or not characteristic parameters change with the observation period (hypothesis 2). In order to estimate a stable price index, an observation period of sufficient length is necessary, but the longer the observation period, the more liable these kinds of structural changes are to occur. The biggest problem is the downward effect on real estate prices due to deterioration pointed out by Bailey et al. (1963)[2], Palmquist (1979)[54], and others. Below, we will show that it is not possible to separate the time effect and age effect in standard repeat sales price indexes and discuss methods for simultaneously estimating the time effect and age effect.

As the transaction interval becomes greater, real estate depreciates and the market value declines. This is explained by McMillen (2003)[47]'s succinct model. The real estate at time t is defined as $p^t = Q^t H^t$. Here, $p^t = \exp(\alpha_t + \gamma_x x)$ is the price per unit of floor space, which varies based on the location (distance from city center) x, α_t is the time effect, and $\gamma_x < 0$. H^t is the real estate floor space, which is produced using land L and capital K^t (a linear homogeneous Cobb-Douglas production function is hypothesized: $H^t = L^{1-\xi} K^{\xi}$). Since real estate materials deteriorate over time, this is defined as $K^t = K[0]\exp(c\tau)$. Here, $K[0]$ is the real estate capital when the building age is 0, $c < 0$ is the capital decrease rate per period, and τ is the building age. Based on the above, the logarithm of the real estate price at time t may be written as follows:

$$\ln p^t = \ln Q^t + \ln H^t = \alpha_t + \gamma_x x + (1-\xi)\ln L + \xi \ln K[0] + \theta\tau$$

Note that $\theta = \xi c < 0$. Therefore, the logarithmic price differential for property n transacted twice at time s and time t is as follows:

$$Y_n = \ln p^t - \ln p^s = \alpha_t - \alpha_s + \theta A_n + \nu_n \tag{77}$$

Here, $A_n = t - s > 0$, and v_n is the error term differential. If taking housing capital deterioration into account, it is necessary to consider not only the time effect difference for the logarithmic price differential but also the age effect difference.

The time effect estimator with the BMN-type repeat sales method is $\hat{\boldsymbol{\alpha}} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\boldsymbol{Y}$, but if the real data generation process is (77), then based on

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'(\boldsymbol{A}\theta + \boldsymbol{v}),$$

one obtains

$$E(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\boldsymbol{A}\theta \tag{78}$$

and as long as it is $\theta < 0$, the BMN-type repeat sales method time effect has a bias. In other words, an age effect is included in the BMN-type time effect.

However, it is not possible to simultaneously estimate the time effect and age effect in (77) in order to distinguish them. If (63) is used, the linear relationship shown below is formed between the age difference for property n and the time dummy difference (Cannaday, Munneke, and Yang, 2005[50]):

$$A_n = t - s = \boldsymbol{D}'_n \boldsymbol{u} = \sum_{u=1}^{T} D_n^u u = \underbrace{D_n^1}_{=0} \cdot 1 + \cdots + \underbrace{D_n^s}_{=-1} \cdot s + \cdots + \underbrace{D_n^t}_{=1} \cdot t + \cdots + \underbrace{D_n^T}_{=0} \cdot T$$

Therefore, the repeat sales regression model with the age effect is

$$\boldsymbol{Y} = \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{A}\theta + \boldsymbol{v} = \mathbf{D}(\boldsymbol{\alpha} + \boldsymbol{u}\theta) + \boldsymbol{v} \tag{79}$$

and it is not possible to distinguish between $\boldsymbol{\alpha}$ and θ. Based on the above, there is a difficult problem with the BMN-type repeat sales method: if the effect of deterioration is ignored, a bias occurs in the price index, and if the effect of deterioration is considered, it cannot be estimated due to multicollinearity.

3.2.2 Age Effect Estimation

The problem of not being able to distinguish the time effect and age effect with the standard repeat sales method has been pointed out by Bailey, Muth, and Nourse (1963)[2], Palmquist (1979)[54], Hill, Knight and Sirmans (1997)[59], Hill, Sirmans and Knight (1999)[43], Chau, Wong, and Yiu (2005)[51], Cannaday, Munneke, and Yang (2005)[50], and others. It is common for the age effect to be ignored in the estimation of repeat sales price indexes.

However, estimating the age effect with the repeat sales method is a very significant issue. First, in economic accounting, as typified by the SNA, estimation of housing stocks is performed, but no consistent method has been established for measuring the depreciation rate. In general, the hedonic method is used for this kind of quality adjustment, but a lot of housing characteristic information is needed to estimate the hedonic function. When collecting this information is difficult, it becomes necessary to perform estimation using another approach. If measurement of the depreciation rate is possible within the framework of repeat sales method estimation, it could be applied in many countries. An additional issue is the elimination of the biases inherent in housing price indexes. There is a strong possibility that biases due to being unable to eliminate the age effect with the repeat sales method will be a serious problem, especially in Asian countries like Hong Kong and Japan where high depreciation rates may be expected.

In an attempt to estimate (21), Palmquist (1979)[54] estimated θ independently of the repeat sales regression equation, then adjusted the time effect to satisfy (21). In order to estimate the age effect, Cannaday, Munneke, and Yang (2005)[50] proposed a multivariate repeat sales model that incorporates a dummy variable for building age instead of a continuous term as in (77). In addition to this, a method has been proposed that performs estimation by disrupting the linear relationship between the time dummy variable and age. Chau, Wong, and Yiu (2005)[51] distinguished the time effect and age effect by hypothesizing non-linearity in the age effect (Box-Cox transformation), and Hill, Knight, and Sirmans (1997)[59] did so by refining Case and Quigley (1991)[8]'s hybrid method (joint hedonic and repeat sales model estimation).

Cannaday, Munneke, and Yang (2005)[50] proposed the following model using an age dummy variable:

$$Y_n = \boldsymbol{D}'_n \boldsymbol{\beta} + \boldsymbol{B}'_n \boldsymbol{\theta} + v_n, \quad n = 1, \ldots, N \tag{80}$$

Here, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_t, \ldots, \beta_T)'$ is the unknown time effect parameter, and $\boldsymbol{B}_n = (B_n^1, \ldots, B_n^j, \ldots, B_n^{J-1})'$ is the dummy variable corresponding to building age, defined as follows:

$$B_n^j = \begin{cases} 1 & j = \tau & \text{Building age at second transaction time} \\ -1 & j = \tau - (t-s) & \text{Building age at first transaction time} \\ 0 & \text{Other} \end{cases}$$

As well, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_j, \ldots, \theta_{J-1})'$ is the age dummy coefficient vector. For the building age dummy, in the case of new construction ($j = 0$), B_n^0 is removed, and in the case of the maximum building age value in the sample ($j = J$), B_n^J is removed. By dropping the time dummy for time 0 D_n^0 and removing B_n^0 and B_n^J, it is possible to avoid multicollinearity. According to Cannaday et al. (2005)[50], the reason for this is that since, in general, building age $[0, J]$ has a broader range than the observation period $[0, T]$ in the data used, it may be considered that the degree of freedom lowering effect is less if the two building age dummies are dropped. Dropping the first and last building age dummies is equivalent to assuming that the price change rate is 0 in this sample range. Now, let us take the average value of the price change rate per year of building age as $\bar{\theta}$. However, since in general it may be considered that $\bar{\theta} < 0$, assuming that the price change rate is 0 for the interval between the first and second transactions $t - s$ means overestimating the price change rate average value $\bar{\theta}$, which leads as a result to underestimating the time effect. Therefore, for the time effect, it is necessary to perform upward correction for $-\bar{\theta}(t-s) > 0$ only, and for the age effect, it is necessary to perform downward correction for $\bar{\theta}(t-s) < 0$ only.

Taking the above into account, if we estimate $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}$ from (80) with the method of least squares, using time s as the baseline year and time t as the comparison year, we can define an age-adjusted multivariate repeat sales price index (AAMRS price index) as follows when the building age for the baseline year is j:

$$I_{s,t,j}^{AAMRS} = \frac{\exp\left[\hat{\beta}_t - \bar{\theta}(t-s) + \theta_{j+(t-s)} + \bar{\theta}(t-s)\right]}{\exp\left(\hat{\beta}_s + \theta_j\right)} = \exp\left(\hat{\beta}_t - \hat{\beta}_s + \theta_{j+(t-s)} - \theta_j\right) \quad (81)$$

Here, for the period $t - s$, the price change rate based on the time effect is $\hat{\beta}_t - \hat{\beta}_s$, and the price change rate based on the age effect is $\theta_{j+(t-s)} - \theta_j$. $\bar{\theta}$ is estimated as

$$\ln p^t = \text{const.} + \theta^t \tau_n^t + \text{error}_n, \quad n = 1, 2, \ldots, N(t)$$

for the sample in each period, and the average value of the change rate in relation to building age is obtained from $\bar{\theta} = \sum_t \theta^t / T$. Here, τ_n^t is the building age of property n at time t (81) adjusts the price change based on deterioration in addition to the time effect price change.

A price index in which building age is controlled as a constant (age-constant multivariate repeat sales price index, or ACMRS price index) can be obtained from the following:

$$I_{s,t,j}^{AAMRS} = \exp\left(\hat{\beta}_t - \hat{\beta}_s - \bar{\theta} t\right) \quad (82)$$

In this paper, this will be referred to as a "pure time index."

When the two price indexes were estimated using four cities (Cleveland, Ohio; Miami, Florida; San Francisco, California; and Champaign, Illinois), the results varied for the AAMRS price index when the initial building age value was set as $j = 1$ and $j = 45$. Compared to the orthodox Case-Shiller price index, the price increase rate was smaller for the recently constructed building ($j = 1$) and the price increase rate was higher for the older building ($j = 45$). In other words, one can see that the price index is important even in terms of what should be the baseline building age level. As well, the ACMRS price index, which separates the age effect included in the orthodox Case-Shiller price index from the time effect and holds the age effect constant, showed that, as predicted with (20), three of the cities advanced at a higher level than with the Case-Shiller price index, and the further one gets from the baseline year, the greater the divergence becomes.

Hill, Knight, and Sirmans (1997)[59] distinguished the time effect and age effect by refining Case and Quigley (1991)[8]'s hybrid method (hedonic and repeat sales method joint model estimation). The hedonic regression model is recognized to be defined as follows:

$$y_n^t \equiv \ln p_n^t = \boldsymbol{z}_n'\boldsymbol{\gamma} + \tau_n^t\theta + \boldsymbol{d}_n'\boldsymbol{\beta} + \varepsilon_n^t \quad (n = 1, 2, \ldots, N(t);\ t = 0, 1, \ldots, T) \tag{83}$$

Here, y_n^t is the logarithm of price p_n^t of property n at time t, $\boldsymbol{z}_n$ is the characteristic vector, τ_n^t is the building age of property n at time t, $\boldsymbol{d}_n$ is the time dummy variable, $(\boldsymbol{\gamma}, \theta, \boldsymbol{\beta})$ is the unknown parameter that should be estimated, and ε_n^t is the error term. θ shows the age effect and $\boldsymbol{\beta}$ the time effect.

From sample $n = 1, \ldots, N(t)$, let us take N_R as a property that is transacted twice. When the building age of property n at the first transaction point s is $\tau - (t - s)$ and the building age at the second transaction point t is τ, the repeat sales regression model may be written as follows:

$$Y_n = A_n\theta + \boldsymbol{D}_n'\boldsymbol{\beta} + v_n \quad (n = 1, 2, \ldots, N_R) \tag{84}$$

Here, $A_n = \tau_n - \{\tau_n - (t - s)\}$ is the differential of the building age at time s and time t. If all samples $N + N_R$ for (83) and (84) are pooled, the following regression model is obtained:

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{z} & \boldsymbol{\tau} & \boldsymbol{d} \\ 0 & \boldsymbol{A} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma} \\ \theta \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{v} \end{pmatrix} \tag{85}$$

Case and Quigley (1991)[8] estimated this with the generalized least squares (GLS) method, assuming the error terms for (8) and (9). However, Hill, Knight, and Sirmans (1997)[59] perform the estimation using the maximum-likelihood method, assuming the AR1 process $\varepsilon_n^t = \rho\varepsilon_n^{t-1} + \nu_n^t$ for the hedonic regression model error term. Here, ρ is the autoregressive coefficient, and a time-homogeneous error term is assumed. If $\rho = 1$, the model is the random walk error term in Case and Shiller, but the parameters must be tested. A distinctive feature of this approach is that, by pooling a hedonic regression model and repeat sales regression model, it disrupts the linear relationship between $\boldsymbol{A}$ and $\boldsymbol{D}$ and makes it possible to estimate the age effect θ.

Hill, Knight, and Sirmans (1997)[59] report that the age effect when hedonic regression model (83) is estimated using OLS or GLS and the age effect of the pooled joint model (85) are roughly the same value, the error term's autocorrelation coefficient is significant, and $\hat{\rho} = 0.54$. In the case of the repeat sales price index when (84) is estimated alone, since a negative age effect is included, it is estimated at a lower value than the hedonic price index when (83) is estimated alone. The price index based on the joint model (85) maximum-likelihood method progresses with a somewhat higher value than the hedonic price index. As well, it is shown that the estimator based on the serial correlation obtained with the maximum-likelihood method is also efficient in Monte Carlo testing.

For the repeat sales method, estimation is not workable unless some kind of hypothesis is included for the error term's heterogeneous variance based on factors other than age, but Hill, Sirmans, and Knight (1999)[43] also performed Monte Carlo testing for the repeat sales regression model, assuming various heterogeneous variances and serial correlations, and showed that a model with an error term that assumes time-homogeneous serial correlation provides results that are preferable to a random walk.

3.3 Sample Selection Bias

3.3.1 Selection Bias Elimination

With a repeat sales index, since the price index is estimated using samples that are transacted repeatedly, it has been pointed out that sample selection bias likely exists. For example, Shimizu, Nishimura, and Watanabe (2010)[57] used a hedonic index and repeat sales index to analyze the difference between the two indexes. The findings they obtained showed that when a hedonic index was estimated using only repeat sales samples, the extent to which price fluctuations lagged behind market turning points became greater as the number of repeat sales increased. They concluded that this suggests the existence of a structural sampling bias in repeat sales samples.

As a hypothesis for explaining whether or not housing will appear on the market as a good to be exchanged, the following condition may be considered: the seller's offer price must exceed his or her reservation price. Gatzlaff and Haurin (1997)[31], (1998)[32] verified that, if changes in the housing market's economic conditions influence the determination of offer prices and reservation prices, there is a possibility that housing samples that are actually sold are not random samples. In other words, actual observed transaction prices depend on the stochastic process that generates offer prices and reservation prices. With that in mind, selection bias is eliminated by applying a two-stage estimation method (Heckit method) based on Heckman (1979)[37].

Since the transaction prices at the first and second sale times are observed as a paired data-set only if the seller's offer price exceeds his or her reservation price, the use of selected samples in analysis cannot be avoided. In Gatzlaff and Haurin (1997)[31], correction of selection bias is performed by applying the Heckit method, using the simplest repeat sales regression model proposed by Bailey et al. (1963)[2] as a base. In order for a property to be sold on the market, the seller's offer price has to exceed his or her reservation price, and a transaction price is observed only when that happens. Therefore, based on the fact that the conditional expected value for the closed hedonic price error distribution is not 0, selection bias occurs in the hedonic price.

In the hedonic regression model presented by Gatzlaff and Haurin (1998)[32], selection bias elimination is performed using the method below. Taking the seller's reservation (logarithmic) price as y_n^{tR} and offer (logarithmic) price as y_n^{tO}, the following hedonic regression model may be written:

$$y_n^{tR} = \boldsymbol{z}_n'\boldsymbol{\gamma}^R + \boldsymbol{d}_n'\boldsymbol{\alpha} + \varepsilon_n^{tR}, \quad n = 1, \ldots, N;\ t = 0, 1, \ldots, T \tag{86}$$

$$y_n^{tO} = \boldsymbol{z}_n'\boldsymbol{\gamma}^O + \boldsymbol{d}_n'\boldsymbol{\alpha} + \varepsilon_n^{tO} \quad n = 1, \ldots, N;\ t = 0, 1, \ldots, T \tag{87}$$

Here, the average of the error terms $\varepsilon_n^{tR}, \varepsilon_n^{tO}$ is 0, and the variance/covariance matrix is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{RR} & \sigma_{RO} \\ \sigma_{RO} & \sigma_{OO} \end{pmatrix} \tag{88}$$

The actual transacted price y_n^t is observed only when the offer price exceeds the reservation price. In other words:

$$y_n^t = \begin{cases} y_n^{tO} & \text{if } y_n^{tO} - y_n^{tR} \geq 0 \\ \text{unobserved} & \text{if } y_n^{tO} - y_n^{tR} < 0 \end{cases} \tag{89}$$

Therefore, the transaction price expected value is:

$$E(y_n^t) = \boldsymbol{z}'_n \boldsymbol{\gamma} + \boldsymbol{d}'_n \boldsymbol{\alpha} + E\left(\varepsilon_n^{tO} \,|\, y_n^{tO} - y_n^{tR} \geq 0\right) \tag{90}$$

Since the error term expected value is not 0 and selection bias occurs, Heckman (1979)[37]'s approach is used to correct this. That is, probit estimation is performed for the selection function that determines whether or not housing is put up for sale in the first step, and OLS estimation using an inverse Mills ratio is performed in the second step.

Gatzlaff and Haurin (1997)[31] expanded the above model to the repeat sales regression method. Here, for time s, let us take S_n^{s*} as a latent variable that represents the choice of whether or not to put housing on sale, where the selection mechanism may be written with the following regression equation:

$$S_n^{s*} = \boldsymbol{W}_n^{s\prime} \boldsymbol{\pi} + \varphi_n^s \tag{91}$$

Here, $\boldsymbol{W}_n^s$ is the characteristic vector including the seller's individual characteristics, housing characteristics, geographic environment, etc., $\boldsymbol{\pi}$ is the unknown parameter, and φ_n^s is the error term. The latent variable S_n^{s*} cannot actually be observed. For the first transaction, since the price is observed only when the offer price surpasses the reservation price – i.e., when $y_n^{1O} - y_n^{1R} \geq 0$ – this is taken as $S_n^1 = 1$. For the second transaction to be observed, the first transaction must actually occur. Therefore, the binary variable expressing this may be defined as follows:

$$S_n^2 = \begin{cases} 1 & \text{if } y_n^{1O} - y_n^{1R} \geq 0 \text{ and if } y_n^{2O} - y_n^{2R} \geq 0 \\ 0 & \text{if } y_n^{1O} - y_n^{1R} \geq 0 \text{ and if } y_n^{2O} - y_n^{2R} < 0 \\ \text{unobserved} & \text{if } y_n^{1O} - y_n^{1R} < 0 \end{cases} \tag{92}$$

Therefore, the first and second prices y_n^1, y_n^2 are observed if $S_n^2 = 1$ ($y_n^{1O} - y_n^{1R} \geq 0$ and $y_n^{2O} - y_n^{2R} \geq 0$) and cannot be observed in other cases. Taking the error terms of the selection function that determines the first and second sales as φ_n^1, φ_n^2 and the error terms of the hedonic regression models as $\varepsilon_n^1, \varepsilon_n^2$, the variance/covariance matrix may be defined as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & 1 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix} \tag{93}$$

Since both the first and second transaction prices y_n^1, y_n^2 are only observed when $S_n^1 = 1$ and $S_n^2 = 1$ are established simultaneously, the hedonic regression model error term expected values, based on the conditional expected values $E\left(y_n^1 \,|\, S_n^1 = 1 \text{ and } S_n^2\right)$ and $E\left(y_n^2 \,|\, S_n^1 = 1 \text{ and } S_n^2\right)$, are:

$$E\left(\varepsilon_n^1 \,|\, S_n^1 = 1 \text{ and } S_n^2\right) = \sigma_{13}\lambda_n^1 + \sigma_{23}\lambda_n^2$$
$$E\left(\varepsilon_n^2 \,|\, S_n^1 = 1 \text{ and } S_n^2\right) = \sigma_{14}\lambda_n^1 + \sigma_{24}\lambda_n^2$$

Here, λ_n^1, λ_n^2 are inverse Mills ratios. Therefore, the repeat sales regression model corrected for sample selection bias may be written as follows:

$$Y_n = \boldsymbol{D}'_n \boldsymbol{\alpha} + (\sigma_{14} - \sigma_{13})\lambda_n^1 + (\sigma_{24} - \sigma_{23})\lambda_n^2 + \eta_n, \quad n = 1, \ldots, N \tag{94}$$

Analysis by Gatzlaff and Haurin (1997)[31] using Miami housing market data showed that a standard repeat sales price index has an upward bias compared to a price index estimated with (36).

3.3.2 Matching Estimation

Properties transacted multiple times are limited even for the housing market as a whole, and with the standard repeat sales method, data transacted only once is not used at all. Therefore, the problem of selection bias discussed in the previous section occurs, and the reduction in sample size also leads to a decrease in estimation efficiency.

McMillen (2012)[48] proposes a price index estimation method using a matching approach to handle this problem. The time effect $\boldsymbol{\alpha}$ in the hedonic regression model corresponds to treatment effects in policy evaluation. As a simple example, the hedonic regression model in the case of the two times $t = 0, 1$ is written as follows:

$$\begin{aligned} y_n^t &= \alpha_0 + \alpha_1 d_n^1 + \boldsymbol{z}'_n \boldsymbol{\gamma} + \varepsilon_n^t \\ &= \alpha_0 + (\alpha_1 - \alpha_0) d_n^1 + \boldsymbol{z}'_n \boldsymbol{\gamma} + \varepsilon_n^t \quad n = 1, \ldots, N(t);\ t = 0, 1 \end{aligned} \tag{95}$$

Here, d_n^1 is a time dummy, where if $t = 1$, then $d_n^1 = 1$; otherwise, $d_n^1 = 0$. Taking the baseline point as $t = 0$, it is possible to measure the price change rate from d_n^1's coefficient estimation value. In the case of the repeat sales method, the price change rate can be obtained by regressing $y_n^2 - y_n^1$ on d_n^1. One can see that the repeat sales method price change rate observes the difference between a representative "post-treatment" value and representative "pre-treatment" value.

Apart from these methods, it is possible to measure the extent to which price changes occur with the average for the whole sample, using the project evaluation method. In other words, it is necessary to obtain the average treatment effect (ATE):

$$ATE = \frac{1}{N(1)} \sum_{n=1}^{N(1)} d_n^1 E(y_n^1 - y_n^0)$$

This shows that the average price change expected value for property transacted at the baseline time $t = 0$ and re-sold at time $t = 1$ is equivalent to ATE. Or, it shows the average difference in the pre-treatment and post-treatment values. To measure the price index with respect to $t = 1, 2, \ldots, T$, it is necessary to obtain the following:

$$ATE = \frac{1}{N(t)} \sum_{n=1}^{N(t)} d_n^t E(y_n^t - y_n^0), \quad t = 1, 2, \ldots, T \tag{96}$$

To approximate the ATE as the "treatment group average treatment effect," the data for all observation times must be randomly sampled data. In the case of data used with the repeat sales method, treatment group data may only be observed at the points when properties are actually sold. Therefore, it is necessary to match data corresponding to the baseline year $t = 0$ control group to $t = 1, 2, \ldots, T$ at each point in time. To perform matching, we first obtain a propensity score after performing logit regression of the time dummy on the characteristic variable used in the hedonic regression model. Next, we create the treatment group matching data for each time based on kernel matching (Heckman, Ichimura, and Todd, 1998[38]), and finally measure the price index by calculating (38).

McMillen (2012)[48] estimated a price index using quarterly data from 1993 to 2008 (approximately 60 quarters) for single-family housing in Chicago. The data size is approximately 169,000 samples, of which 52,000 are repeat sales data transacted at least twice. The baseline point is Q1 1993, and the control group data consists of 1,651 samples. Due to the nature of repeat sales data, there are few samples at the initial observation time and even fewer samples at the final observation time. The number of matched samples, however, is roughly the same at each point. The total number of matched samples for the period from Q1 1993 to Q4 2008 is 102,000, which exceeds the repeat sales data. When price indexes were estimated based on the hedonic approach using the initial 169,000 samples and 102,000 matched samples, there was almost no difference between the two. In other words, when it comes to the matching estimator, this shows that a hedonic approach-based estimator is extremely robust.

The matching estimator clearly differs from the simple price change average. The price index based on the average value for each period using all 169,000 samples is easily influenced by values that are outliers from the distribution. In the case of Chicago, since the variation in the 2005 data is greater (in particular, the left side is flat) than that of the 1995 data, it is shown that the price index based on the average value for each time is pulled downward.

The repeat sales method of estimating price indexes could be called an extreme version of matching method-based estimation. Matching housing transaction data that is not necessarily identical but is similar has a number of advantages compared to price index estimation with either the repeat sales method or hedonic method. Compared to the standard repeat sales method, the general matching method dramatically increases the sample size and the likelihood of obtaining more efficient estimation values. For example, even in cases where there are few samples (small region, short observation period, etc.), it may make it possible to create a price index. This means that matching-based estimation could be a useful price index estimation method.

3.4 Characteristics, Advantages, and Disadvantages of Repeat Sales Indexes

This chapter has presented an overview of the repeat sales method and discussed it, with a focus on what kind of problems occur. Since the repeat sales method involves price comparison of the same property, if there is no change in characteristics or characteristic prices, the problem of underestimation bias that occurs with the hedonic method is eliminated. As well, since the estimation method is simple, it has the benefits of being an approach with high reproducibility and estimation efficiency.

In order to create a more stable price index, it is necessary to observe price data over an extended period. However, when the observation period becomes longer, aggregation bias occurs due to changes in characteristics and characteristic values for the same property.

Since the price of housing changes due to deterioration and investment in renovations (housing age effect), an age effect is included in the time effect in the standard repeat sales method. However, since a perfectly linear relationship exists between the time dummy and the variable indicating the transaction interval, it is not possible to distinguish the time effect and age effect in the standard method. What have been proposed to date are methods that intentionally disrupt the linear relationship between the time dummy and transaction interval variable and methods that extrapolate the price index using exogenous data.

As well, since only property transacted multiple times is selected for use with the repeat sales method, the sample size is reduced and the occurrence of selection bias in the sample is a concern. If changes in housing market economic conditions influence the determination of the offer price and reservation price, since the transactions at the times of the first and second sales are observed as a paired data-set only when the seller's offer price exceeds the reservation price, the use of selected samples in analysis cannot be avoided. In this case, the traditional method of correcting bias by estimating a selection function has been proposed.

Properties transacted multiple times are limited even for the housing market as a whole, and with the standard repeat sales method, data transacted only once is not used at all. It has been shown that the matching method is useful in improving this point. Using data transacted only once as re-sold data based on the matching method increases the sample size dramatically and the likelihood of obtained more efficient estimation values. This may, for example, make it possible to create a price index even in cases where there are few samples (small region, short observation period, etc.).

To summarize the above points, the advantages include:

- Since the index is created by comparing prices of repeatedly transacted properties at different points in time, there is no need for information relating to the property characteristics.
- The problem of omitted variable bias that occurs with the hedonic method is avoided.
- The estimation method is simple and there is a high level of reproducibility.
- Even in the case of strong heterogeneous property, the probability of estimating index is high.
- Due to simple concept, it is easy to explain to users.

The disadvantages include:

- Since the price index is estimated using only information for properties transacted at least twice (information for properties transacted only once is discarded), this method is inefficient. As a result, its use is difficult in countries or regions where liquidity is low, and it often becomes difficult to estimate indexes restricted to certain regions or property uses.
- Since the depreciation that accompanies the aging of the building between the two transaction times is ignored, there is a downward bias if this is not controlled for.
- If investment in renovations is made between the two transaction times, there is an upward bias if this is not controlled for.
- Depending on the database composition, it may be cost-intensive to identify transactions involving the same property (there are quite a few countries where it is difficult to identify transactions involving the same property).
- It is impossible to create separate indexes for land and buildings.
- When new transaction price information is generated, the data – including even past series – changes, so it is not possible to produce definite values.

4 Price Indexes Based on Property Appraisal Prices

4.1 Property Appraisal Price Indexes

If the property market has few transactions (i.e., it is thin) and property is strongly heterogeneous, price surveys are conducted by property appraisal experts. In addition, in the many countries with property taxes, there are quite a few that use property assessment values for the purpose of tax assessment.

Moreover, in recent years, with the dramatic growth in the property investment market, it has become possible to obtain property appraisal prices that are periodically surveyed for the purpose of measuring the performance of investment properties. In light of this, efforts have been made to create property prices indexes using property appraisal prices.

In particular, when attempting to capture the movements of markets that are strongly heterogeneous with few transactions, using property appraisal prices may be a valuable method of capturing changes.

However, it has been pointed out that there are valuation error, lagging, and smoothing problems surrounding property appraisal prices. The first problem occurs because property appraisal prices are determined based on the judgment of property appraisal experts, so there is a certain degree of error in the price determination absolute value. The second problem occurs because the information property appraisers are able to use in price determination is past information, so there is a certain lag in price determination. The final problem, which is related to the first and second problems, occurs because not only is there a strong possibility of misjudging market turning points, but changes also undergo smoothing, so price changes occur only gradually.

Furthermore, property appraisal systems differ by country, so there are cases where the definition of the price obtained by property appraisers also varies. Furthermore, assessed values for the purpose of tax assessment differ from normal property appraisal values, and since they are assessed values, there is an even stronger possibility that they do not properly capture market changes.

While property appraisal price information is a valuable information source for markets with few transactions, and the possibility of creating an index using this information exists, sufficient care must be taken with regard to its biases.

4.2 Hedonic Method Based on Pooling of Property Appraisal Prices and Transaction Prices

When attempting to estimate hedonic price indexes using transaction price information, one faces cases where index estimation is difficult due to a lack of such information. In addition, as mentioned previously, since valuation error and smoothing problems exist with property price information, it is known to have certain biases. In order to control for these biases and compensate for insufficient transaction price information, the following has been proposed: attempting to estimate price indexes by pooling property appraisal price information and transaction price information, then using the hedonic method.

Since assessed values obtained for tax purposes incorporate various factors that are likely included in transaction price information, they are helpful in explaining transaction prices. The regression equation using assessed values is written as follows:

$$\ln p_n^t = \alpha_t + \varsigma \ln a_n^t + \varepsilon_n^t \tag{97}$$

Here, p_n^t is the transaction price and a_n^t is the assessed value. Note that there is no guarantee that the true property market value can be assessed correctly, so there is always an error in the assessed value. For example, taking the true property market value as V_n^t, the assessed value a_n^t may be an observation value accompanied by a probability error, as follows:

$$\ln a_n^t = \ln V_n^t + \eta_n^t \tag{98}$$

In other words, the assessed value is the true property market value with the probability error η_n^t added to it. If the assessed value in (99) is taken as a proxy for the true property market value, since $\ln V_n^t = \ln a_n^t - \eta_n^t$ based on (98), (99) may be rewritten as follows:

$$\ln p_n^t = \alpha_t + \varsigma\left(\ln a_n^t - \eta_n^t\right) + \varepsilon_n^t = \alpha_t + \varsigma \ln a_n^t + \left(\varepsilon_n^t - \varsigma\eta_n^t\right) \tag{99}$$

Since the explanatory variable $\ln a_n^t$ is clearly correlated to the error term, the coefficient's least-squares estimator has a bias.

This estimation method is currently being researched and developed by the European Central Bank with the aim of applying it in practice.

4.3 The SPAR Method

Since there is also a lack of price information for properties transacted multiple times when estimating repeat sales indexes, one may be faced with the problem of being unable to estimate the price index. In light of this, along with the method of artificially increasing the number of repeat sales using the previously mentioned matching method, an estimation method known as the SPAR (sale price appraisal ratio) method, which obtains the first transaction price with the property appraisal price, has been proposed and applied in practice.

The sale price of property n at comparison point t is taken as $p_n^t\,(n = 1, 2, \ldots, N(t))$. In addition, the appraisal price of said property at the baseline point 0 is taken as a_n^0 $(n = 1, 2, \ldots, N(0))$. In this case, the sale-appraisal price ratio is p_n^t/a_n^0. If all quantities are standardized as 1, the appraisal price-based arithmetic average price index may be defined as follows:

$$P_{AP}^{0t} = \frac{\sum_{n=1}^{N(t)} p_n^t}{\sum_{n=1}^{N(t)} a_n^0} = \sum_{n=1}^{N(t)} w_n^0(t)\left(\frac{p_n^t}{a_n^0}\right) \tag{100}$$

Here, $w_n^0(t)$ in the second formula on the right side of (100) is the weight based on the appraisal price, and $w_n^0(t) = a_n^0 \Big/ \sum_{n=1}^{N(t)} a_n^0$. Since this weight is defined by the quantity (standardized as 1) of sample $N(t)$ at the comparison point, $w_n^0(t)$ is the expenditure weight calculated with the baseline point price and comparison point quantity. Therefore, since (100) is the weighted average based on $w_n^0(t)$ in the sale-appraisal price ratio p_n^t/a_n^0, one can see that it is a Paasche-type index. Note that in general, the baseline point sample size $N(0)$ and comparison point sample size $N(t)$ are not equivalent.

The problem with (100) is that it takes the appraisal value as the baseline point price. Since the sale price is not used, the price index is not 1 at the baseline point. The arithmetic average sales price appraisal ratio method (arithmetic method) index overcomes this problem by dividing by the baseline point sale-appraisal price ratio, as follows:

$$P_{SPAR}^{0t} = \frac{\sum_{n=1}^{N(t)} p_n^t}{\sum_{n=1}^{N(t)} a_n^0}\left(\frac{\sum_{n=1}^{N(0)} p_n^0}{\sum_{n=1}^{N(0)} a_n^0}\right)^{-1} = \frac{\sum_{n=1}^{N(t)} p_n^t/N(t)}{\sum_{n=1}^{N(0)} p_n^0/N(0)}\left(\frac{\sum_{n=1}^{N(0)} a_n^0/N(0)}{\sum_{n=1}^{N(t)} a_n^0/N(t)}\right) \tag{101}$$

(101) is a reciprocal multiplication of the sale price arithmetic average ratio and appraisal price arithmetic average ratio. The reciprocal of the appraisal price arithmetic average plays a role in adjusting structural changes that occur from the baseline point to the comparison point.

4.4 Characteristics, Advantages, and Disadvantages of Property Appraisal Price Indexes

Property appraisal price information is, needless to say, an extremely important source of information in the estimation of property price indexes. In particular, in regions where there are few transactions and markets which are strongly heterogeneous, such as the logistics facility, hotel, or hospital markets, there are quite a few cases where one has to rely on property appraisal price information.

In light of this, not only are there price indexes that make direct use of property appraisal prices, but many inventive approaches have also been developed, such as methods like the SPAR method that correct the repeat sales method by using property appraisal prices and methods that perform estimation by combining property appraisal prices and transaction prices in hedonic method estimation. Their respective advantages and disadvantages are outlined below. First, the SPAR method's advantages include:

- It preserves the advantages of the repeat sales method.
- Since it enables the use of more information than the repeat sales method, it is highly efficient.
- Since it is a method based on traditional index theory, it is easy to understand, and the estimation method is simple, it has a high level of reproducibility.

Its disadvantages include:

- It inherits the disadvantages of the repeat sales method.
- Since the initial transaction is obtained with the property appraisal price, it is affected by the valuation error and smoothing problems with property appraisal prices.
- The quality adjustment issues surrounding property price indexes.

In the case of estimation using property appraisal price information and transaction price information with the hedonic method, while it artificially increases the number of samples and increases the efficiency when estimating the hedonic function, numerous problems remain in terms of estimation theory, such as how to set the probability that transactions will occur.

5 How Should Property Price Indexes Be Estimated?

How should property price indexes be estimated?

When estimating a property price index, the estimation method varies considerably based on the limitations of available information. If no such limitations exist, the hedonic method has an advantage when one considers the underlying economic theory, the consistency with other types of economic statistics, its application in the System of National Accounts, and so forth. However, in reality, while it may be viable for the housing market, where the transaction quantity is relatively large and quality is relatively high, or even for the office market when it comes to commercial property, in markets that are strongly heterogeneous, there are quite a few cases where it is difficult to apply the hedonic method.

The repeat sales method is effective when there is a sufficient quantity of transactions, even if the market is strongly heterogeneous. However, in markets where the number of transactions is limited, application of the repeat sales method is also difficult.

In such cases, it may be possible to create indexes using property appraisal price information.

There are various possibilities, such as the SPAR method, estimation based on the hedonic method using property appraisal prices and transaction prices, and appraisal price indexes that use property appraisal prices as is.

However, when there are few transactions, one faces the problem of how property appraisers determine property appraisal prices and, in light of that, how reliable the determined property appraisal prices are.

Furthermore, in cases where there is a lack of property transaction price information, there is further scope to consider creating indexes using property revenue information and so on.

Going forward, in an attempt to properly capture property market trends, it is likely that multiple indexes will be created by combining various sources of information with appropriate estimation methods for those sources.

References

[1] Anglin, P. M. and R. Gencay (1996), "Semiparametric Estimation of Hedonic Price Function," Journal of Applied Econometrics, Vol. 11, No. 6, pp. 633–648.

[2] Bailey, M. J., R. F. Muth and H. O. Nourse (1963), "A Regression Model for Real Estate Price Index Construction," Journal of American Statistical Association, Vol. 58, pp. 933–942.

[3] Barten, A. P (1964), "Consumer Demand Functions under Conditions of Almost Additive Preferences," Econometrica, Vol. 32, No. 1/2, pp. 1–38.

[4] Berndt, E. R., Z. Griliches and N. J. Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990s?," Journal of Econometrics Vol. 68, pp. 243–268.

[5] Box, G. E. P. and D. R. Cox (1964), "Ananalysis of Transformations," Journal of the Royal Statistical Society. SeriesB (Methodological), Vol. 26, No. 2, pp. 211–252, .

[6] Brown, J. N. and H. S. Rosen (1982), "On the Estimation of Structural Hedonic Price Models," Econometrica, Vol. 50, No. 3, pp. 765–768.

[7] Case, B., H. O. Pollakowski and S. M. Wachter (1991), "On Choosing Among House Price Index Methodologies," Real Estate Economics Vol. 19(3), pp. 286–307.

[8] Case, B. and J. M. Quigley (1991), "The Dynamics of Real Estate Prices," Review of Economics and Statistics, Vol. 73, No. 1, pp.50–58.

[9] Case, K. E. and R. J. Shiller (1987), "Prices of Single Family Homes Since 1970: New Indexes for Four cities," New England Economic Review, Vol. (Sept./Oct.), pp. 45–56.

[10] Case, K. E. and R. J. Shiller (1989), "The Effciency of the Market for Single-Family Homes," The American Economic review, Vol. 79, No. 1, pp.125–137.

[11] Cassel, E. and R. Mendelsohn (1985), "The Choice of Functional Forms for Hedonic Price Equations: Comment," Journal of Urban Economics, Vol. 18, No. 2, pp. 135–142.

[12] Clayton, J., D. Geltner and S. W. Hamilton (2001), "Smoothing in Commercial Property Valuations: Evidence from Individual Appraisals,"" Real Estate Economics, Vol. 29, issue 3, pp. 337–360.

[13] Cropper, M. L., L. B. Deck and K. E. McConnell (1988), "On the Choice of Functional Form for Hedonic Price Functions," The Review of Economics and Statistics, Vol. 70, No. 4, pp. 668–675.

[14] Court, A. T (1939), "Hedonic Price Indexes with Automotive Examples," in The Dynamics of Automobile Demand, General Motors, New York.

[15] Devaney, S. and R. M. Diaz (2010), "Transaction Based Indices for the UK Commercial Property Market: Exploration and Evaluation Using IPD Data," Discussion Paper 2010-

02, University of Aberdeen Buisiness school, pp. 1-19.
[16] de Haan, J. (2004), ”Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement,” Journal of Economic and Social Measurement, Vol. 29(4), pp. 427–443.
[17] de Haan, J. (2010), ”Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-Pricing methods,” Jahrbücher für Nationalökonomie und Statistik, Vol. 230(6), pp. 772–791.
[18] Diamond, D. Jr. and B. A. Smith (1985), ”Simultaneity in the Market for Housing Characteristics,” Journal of Urban Economics, Vol. 17, No. 3, pp. 280–292.
[19] Diewert, W. E. (1971), ”An Application of the Shepherd Duality Theorem: A Generalized Leontief Production Function,” Journal of Political Economy, Vol. 79, No. 3, pp. 481–507.
[20] Diewert, W. E. (1973), ”Functional Forms for Profit and Transformation functions,” Journal of Economic Theory, Vol. 6, No. 3, pp. 284–316.
[21] Diewert, W. E. (1976), ”Exact and Superlative Index Numbers,” Journal of Econometrics, Vol. 4, No. 2, pp. 114–145.
[22] Diewert, W. E. (1983), ”The Theory of the Output Price Index and the Measurement of Real Output Change,” in Price Level Measurement, W.E. Diewert and C.Montmarquette (eds.), Ottawa: Statistics Canada, pp. 1049–1113 .
[23] Diewert, W. E. (1997), ”Commentary on Mathew D. Shapiro and David W. Wilcox: Alternative Strategies for Aggregating Price in the CPI,” The Federal Reserve Bank of St. Louis Review, Vol. 79:3, pp. 127–137.
[24] Diewert, W. E. (2002), ”Hedonic Producer Price Indexes and Quality Adjustment,” Discussion Paper 02-14, Department of Economics, University of British Columbia, pp. 1–11.
[25] Diewert, W.E. (2003), ”Hedonic Regressions: A Review of Some Unresolved Issues,” Mimeo, Department of Economics, University of British Colombia.
[26] Diewert, W. E. (2007), ”The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions,” Discussion Paper 001-07, Department of Economics, University of British Columbia, pp. 1–33.
[27] Diewert, W. E., S. Heravi and M. Silver (2007), ”Hedonic Imputation Versus Time Dummy Hedonic Indexes,” Discussion Paper 007-07, Department of Economics, University of British Columbia, pp. 1–29.
[28] Dulberger, E.R. (1989), ”The Application of a Hedonic Model to a Quality-Adjusted Price Index for Computer Processors,” In D. W. Jorgenson and R. Landau (eds.), Technology and Capital Formation (pp. 37–75), Cambridge, MA: MIT Press.
[29] Ekeland, I., J. J. Heckman and L. Nesheim (2004), ”Identification and Estimation of Hedonic Models,” Journal of Political Economy, Vol. 112, No. S1, pp. S60–S109.
[30] Eurostat (2013) Handbook on Residential Property Price Indices (RPPIs), Methodologies & Working papers, 2013 edition.
[31] Gatzlaff, D. H. and D. R. Haurin (1997), ”Sample Selection Bias and Repeat-Sales Index Estimates,” Journal of Real Estate Finance and Economics, Vol. 14, No. 1/2, pp. 33–50.
[32] Gatzlaff, D. H. and D. R. Haurin (1998), ”Sample Selection and Biases in Local House Value Indices,” Journal of Urban Economics, Vol. 43, No. 2, pp. 199–222.
[33] Griliches, Z (1961), ”Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change,” In G. Stigler (chairman), The Price Statistics of the Federal Government, Washington D.C.: Government Printing Office.
[34] Griliches, Z. (1967), ”Hedonic Price Indexes Revisited: A Note on the State of the Art,” Proceedings of the Business and Economics Section of the American Statistical Association, pp. 332–334.
[35] Griliches, Z (1971), ”Introduction: Hedonic Price Indexes Revisited,” In Z. Griliches (ed.) Price Indexes and Quality Change, (pp. 3–15), Cambridge MA: Harvard University Press.

[36] Halvorsen, R. and H. O. Pollakowski (1981), "Choice of Functional Form for Hedonic Price Equations," Journal of Urban Economics, Vol. 10, No. 1, pp. 37–49.
[37] Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, No. 1, pp. 153–161.
[38] Heckman, J. J., H. Ichimura and P. Todd (1998), "Matching as an Econometric Evaluation Estimator," Review of Economic Studies, Vol. 65, No. 2, pp. 261–294.
[39] Heckman, J. J., R. L. Matzkin and L. Nesheim (2010), "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," Econometrica, Vol. 78, No. 5, pp. 1569–1591.
[40] Hill, R. C., J. R. Knight and C. F. Sirmans (1993), "Estimation of Hedonic Housing Price Models Using non Sample Information: A Montecarlo Study," Journal of Urban Economics, Vol. 34, No. 3, pp. 319–346.
[41] Dombrow, J., J. R. Knight and C. F. Sirmans (1997), "Aggregation Bias in Repeat-Sales Indices," The Journal of Real Estate Finance and Economics, Vol. 14 (1-2), pp. 75-88.
[42] Hill R. C., J. R. Knight and C. F. Sirmans (1997), "Estimating Capital Asset Price Indexes," The Review of Economics and Statistics, Vol. 79, No. 2, pp. 226–233.
[43] Hill R. C., C. F. Sirmans and J. R. Knight (1999), "A Random Walk Down Main Street?," Regional Science and Urban Economics, Vol. 29, No. 1, pp. 89–103.
[44] Lancaster, K. J. (1996), "A New Approach to Consumer Theory," Journal of Political Economy, Vol. 74, No. 2, pp. 132–157.
[45] Linneman, P. (1980), "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," Journal of Urban Economics, Vol. 8, No. 1, pp. 47–68.
[46] Malpezzi, S. (2003), "Hedonic Pricing Models: a Selective and Applied Review," In A. O'Sullivan and K. Gibb (eds.), Housing Economics: Essays in Honor of Duncan Maclennan, (pp. 67–89). Blackwell: Malder, MA.
[47] McMillen, D. P. (2003), "The Return of Centralization to Chicago: Using Repeat Sales to Identify Changes in House Price Distance Gradients," Regional Science and Urban Economics, Vol. 33, No. 3, pp. 287–304.
[48] McMillen, D. P. (2012), "Repeat Sales as a Matching Estimator," Real Estate Economics, Vol. 40, No. 4, pp. 745–773.
[49] Mendelsohn, R. (1985), "Identifying Structural Equations with Single Market Data," The Review of Economics and Statistics, Vol. 67, No. 3, pp. 525–529.
[50] Cannaday, R. E., H. J. Munneke and T. T. Yang (2005), "A Multivariate Repeat-Sales Model for Estimating House Price Indices Journal of Urban Economics, Vol. 57, No. 2, pp. 320–342.
[51] Chau, K.W., S. K. Wong and C.U. Yiu (2005), "Adjusting for Non-linear Age Effects in the Repeat Sales Index," Journal of Real Estate Finance and Economics, Vol. 31, No. 2, pp. 137–153.
[52] Pace, R. K (1998), "Appraisal Using Generalized Additive Models," Journal of Real Estate Research, Vol. 15, No. 1/2, pp. 77–99.
[53] Pace, R. K (1995), "Parametric, Semiparametric, and Nonparametric Estimation of Characteristic Values with in Mass Assessment and Hedonic Pricing Models," Journal of Real Estate Finance and Economics, Vol. 11, No. 3, pp. 195–217.
[54] Palmquist, R. B. (1979), "Hedonic Price Depreciation Indexes for Residential Housing: A Comment," Journal of Urban Economics, Vol. 6, No. 2, pp. 267–271.
[55] Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, Vol. 82, No. 1, pp. 34–55.
[56] Sheppard, S. (1999), "Hedonic Analysis of Housing Markets," In Handbook of Regional and Urban Economics Vol.3, Cheshire, P. C. and E. S. Mills (eds.), chapter 41, pp. 1595–

1635.

[57] Shimizu, C., K. G. Nishimura and T. Watanabe (2010), "House Prices in Tokyo –A Comparison of Repeat-sales and Hedonic Measures–," Journal of Economics and Statistics, Vol. 230 (6), pp.792–813.

[58] Silver, M. and S. Heravi (2007), "The Difference Hedonic Imputation Indexes and Time Dummy Hedonic Indexes," Journal of Business and Economic Statistics, vol. 25:2, pp. 239–246.

[59] Witte, A. D., H. Sumka and J. Erekson (1979), "An Estimate of a Structural Hedonic Price Model of the Housing Market: Anapplication of Rosen's Theory of Implicit Markets," Econometrica, Vol. 47, No. 5, pp. 1151–1172.