Constructing a Building Price Index Using Big Data from the Administrative Sector

A Method for Estimating an Output-type Building Price Index Using Individual Data from the Statistics of Building Starts

By

Masahiro Higo (The University of Tokyo)
Yumi Saita (Osaka University)
Chihiro Shimizu (Nihon University and The University of Tokyo)
Yuta Tachi (Reitaku University)

October 2021

CREPE DISCUSSION PAPER NO. 112

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)
THE UNIVERSITY OF TOKYO
http://www.crepe.e.u-tokyo.ac.jp/

Constructing a Building Price Index Using Big Data from the Administrative Sector[*]

A Method for Estimating an Output-type Building Price Index Using Individual Data from the Statistics of Building Starts

Masahiro Higo[†] Yumi Saita[‡] Chihiro Shimizu[§] Yuta Tachi[**]

Abstract

Improving the accuracy of deflators is crucial for measuring real GDP and growth rates. However, construction prices are often difficult to measure. This study uses the stratification and hedonic methods to estimate price indices. The estimated indices are based on the actual transaction prices of buildings (contract prices) obtained from the Statistics on Construction Starts survey information and big data from the administrative sector in Japan.

Compared with the construction cost deflator, calculated by compounding input costs, the estimated output price indices show higher rates of increase during the economic expansion phase after 2013. This suggests that the profit surge in the construction sector observed in that period is not fully reflected in the construction cost deflator. Furthermore, the difference between the two "output-type" indices obtained by stratification and hedonic methods tends to zero when the estimation methods are precisely configured.

Keywords: building price index, stratification method, hedonic method, Japan, big data

JEL classification: C20, C23, C43

[*] We thank Erwin Diewert, Paul Schreyer, Marshall Reinsdolf, Niall O'Hanlon, Kate Burnett-Isaacs, Koji Nomura and Rhys Lewis for their advice and materials. This study was conducted in the office of the Statistics Committee. The authors also thank the Ministry of Land, Infrastructure, Transport, and Tourism for providing the data. This work gratefully acknowledges the support received from JSPS 20H00082.

[†] The University of Tokyo (e-mail: masahiro.higo@e.u-tokyo.ac.jp)

[‡] Osaka University (e-mail: saita@osipp.osaka-u.ac.jp)

[§] Nihon University & The University of Tokyo (e-mail: cshimizu@csis.u-tokyo.ac.jp)

[**] Reitaku University (e-mail: tachiyuta@gmail.com)

1. Introduction

Improving the accuracy of deflators (price indices) is crucial for correctly measuring real GDP and growth rates. However, price indices often deviate from deflators that meet the System of National Accounts (SNA) international standards due to difficult access to price information and technical limitations in indexing. This issue is particularly evident in the construction industry, as the weight of construction investment in a country's GDP is typically large. Hence, improving the accuracy of construction price indices is fundamental for statistical departments in various countries.[1]

In Japan, there is no output-type construction price index based on actual transaction prices (contract prices) of buildings and civil engineering structures. The SNA's deflator uses "input-cost" prices as an alternative price index based on the intermediate input and labor costs required for construction activities. However, the resulting construction deflator does not reflect changes in the profit margins of the construction sector, possibly causing errors in the real value of construction investment. Some foreign countries have adopted output-type construction price indices that directly measure output prices (contract prices), namely actual transaction prices of buildings and civil engineering structures, instead of the conventional input-cost price indices.

In this study, we attempt to create an output-type building price index directly measuring output prices by using survey data from the Statistics on Building Starts, based on the Notification of Building Construction that must be submitted in Japan when constructing a building. To the best of our knowledge, this is the first study to construct a price index using these data. We propose a new approach that utilizes large-scale information ("big data") collected by administrative agencies. This approach may be considered entirely novel, not only in Japan but also in other countries.

The remainder of this paper is organized as follows. Section 2 summarizes the limits of the current input-cost construction price indices and provides an overview of three common approaches to creating output-type construction price indices. Section 3 provides an overview of the information obtained from the Statistics of Construction

---

[1] In Japan, the Statistics Commission, the command post for statistics, together with the ministries and agencies that prepare statistics, has been working on reforms required by the government's Council on Economic and Fiscal Policy and the Council for the Promotion of Statistical Reform. The "Basic Plan for the Development of Official Statistics" (the 3rd Basic Plan), approved by the Cabinet in March 2018, states that "from the perspective of improving the accuracy of real values (omission), research and empirical studies for the use of market-based prices for construction and retail services (margins) will be promoted." The report recognizes that improving the accuracy of the deflator (price index) in realizing nominal values will substantially improve the accuracy of GDP estimates.

Starts, explaining how the stratification and hedonic methods are used to create the proposed price index. Section 4 reports the estimation results of the output-type building price indices obtained through the stratification and hedonic methods, comparing them with the input-cost construction cost deflator. Additional analyses and results are then proposed and discussed. Section 5 summarizes the study's findings and outlines the remaining issues for future research.

2. Output-Type Construction Price Indices

2-1. Limits of the Current Input-Cost Construction Price Index

2-1-1. Factors of the use of input-cost price indices in the construction sector

The SNA calculates real amounts by dividing nominal amounts, such as the production value, by a deflator.[2] A deflator is a price index calculated by continuously surveying the prices of goods and services with constant quality through repeated transactions, indexing the prices of each product, so that the base point is equal to 100, and weighting the price indices of individual products using weights corresponding to the transaction amounts. In Japan's SNA, the Consumer Price Index produced by the Statistics Bureau of the Ministry of Internal Affairs and Communications; the Corporate Goods Price Index and the Services Producer Price Index produced by the Bank of Japan; and the Agricultural Price Index produced by the Ministry of Agriculture, Forestry and Fisheries are typically used.

However, in the construction sector, there is no market transaction price-based construction price index. In the case of goods and services covered by the consumer price index and the corporate goods and services price index, goods and services of the same quality are exchanged repeatedly; hence, a constant-quality price index may be created by continuously examining price trends. In contrast, buildings and civil engineering structures are custom-made products, and those of the same quality are not traded repeatedly. Therefore, it is not possible to continuously survey the market transaction prices (contract prices) of buildings and civil engineering structures of the same quality.

The SNA has developed an alternative, input-cost construction price index based on the input costs of buildings and civil engineering structures and uses it as a deflator. The

[2] See Cabinet Office (2021) for detailed instructions on how to create an index.

Bank of Japan's Corporate Goods Price Index and the Services Producer Price Index are employed for the materials used for assessing construction activities (intermediate inputs), and the Ministry of Health, Labor, and Welfare's Monthly Labor Survey is used to derive per capita wages in the construction industry, used for assessing labor costs (employer compensation). In addition, the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT)[3] produces the Construction Cost Deflator, one of Japan's most used input-cost construction price indices.

2-1-2. Factors causing bias in the input-cost construction price index

The current input-cost construction price indices, such as the SNA construction deflator, have two major limitations.

First, these indices only cover the intermediate inputs and compensations of employees.[4] Therefore, price changes due to a variation in the profit of the construction industry, not considered by the current input-cost construction price index, are assumed to be equal to those of intermediate inputs and employees' compensation. If these price changes are different, the resulting price index is substantially biased, as the coverage of the input-cost price index changes over time.

Second, the Monthly Labor Survey only reports the average per capita wages of construction workers, used as price data to assess labor costs; hence, it does not consider changes in the quality of labor, such as age, length of service, education, and employment status. Employer compensation accounts for more than 30% of the construction industry's output (35% in 2019). As Fukao et al. (2017) point out, its impact is significant and likely to influence changes in building quality. For example, according to the Japan Industrial Productivity (JIP) Database 2021, the quality of labor in the construction industry (the average of building and civil engineering) has improved by 12% from 1994 to 2018.[5]

2-2. How to Create an Output-Type Construction Price Index

---

[3] For information on how to create a Construction Cost Deflator, see Ministry of Land, Infrastructure, Transport and Tourism (2021).

[4] Intermediate inputs and employer compensation accounted for 90% of the output of construction in 2019. The profit generated by construction activities (operating surplus and mixed income), fixed capital depletion, and taxes (taxes imposed on "produced and imported goods" minus "subsidies"), which accounted for the remaining 10%, were not covered. The profit share of the construction sector, which is not covered, often causes errors in the "input cost" price index as it fluctuates significantly compared to input costs due to changes in the environment in which construction companies receive orders.

[5] The Japan Institute for Labour Policy and Training's "Youthful Labour Statistics 2020" estimates that the rate of increase in simple average wages and average wages with fixed attributes in the construction industry from the Ministry of Health, Labor, and Welfare's "Basic Survey on Wage Structure" shows a 13% increase from 1994 to 2019 due to improvements in the quality of work. The gap is 13% over 25 years.

To overcome insufficient coverage of the price indices and wage data with no fixed quality, it is necessary to create a price index based on market transaction prices, namely, an output-type index reflecting the output prices of buildings and civil engineering structures. The OECD manual "Sources And Methods Construction Price Indices" introduces the following six methods to construct a price index:

(1) Model price method
(2) Quoted prices method
(3) List prices method
(4) Matched models method
(5) Stratification method
(6) Hedonic method

Among these, the methods typically adopted in overseas countries are (1) the model price approach, (5) the per square meter (stratified) approach, and (6) the econometric (hedonic) approach. In this section, we summarize their characteristics.

(1) Model price method

The model price method is a compiling method in which "models" of typical buildings and civil engineering structures are obtained by adding up the hypothetical prices of construction materials, labor, machinery, and equipment for each component, and adding the assumed profit of the construction company. This method has been used in the U.S., Canada, Germany, among others. In Japan, this method has been adopted for some items in the Corporate Goods Price Index and the Services Producer Price Index; however, it has not yet been adopted for the construction cost deflator and the SNA deflator.

For example, in the U.S., the Bureau of Labor Statistics purchases data on past construction projects from construction cost estimating companies, selecting representative construction projects for each region (Northeast, Midwest, South, and West) to construct a "building model.” The construction contract price is calculated by adding all the "assemblies" required to obtain the building model and all the "components" required for the assemblies. In Canada, Statistics Canada designates a representative newly constructed building and asks construction companies the price change of each production factor from the previous quarter. In Germany, the Federal

Statistical Office of Germany directly obtains the unit price of constructions, equivalent to the output price, from the surveyed companies.

However, this method has some limitations. First, the price indices may not reflect the transaction prices traded in actual markets if some of the production factors in the building or construction models are not representative. Second, as the modeled prices are not actual transaction prices, they may not accurately reflect the market situation. Third, and most importantly, the cost of producing these statistics is high. It is often necessary to hire experts for designing a standard model. The burden on construction companies reporting hypothetical estimated prices is also high, and the collection of markup rate data is difficult. Furthermore, quality adjustment of labor costs and estimation of user costs also involve difficulties in terms of data collection and analysis.[6]

(2) Stratification method

This method stratifies price data into multiple attributes, which have a large impact on prices (use, structure, construction method, building method, and region, among others), creating a price index based on the average price for each subdivided stratum.[7] In this method, buildings and civil engineering structures with different qualities are regarded as having the same quality for price data belonging to the same stratum. This approach is commonly employed in existing price indices.

In the stratification method, once the subdivision rules have been determined, the burden of producing price indices remains constant as that of tabulating ordinary statistical surveys, and price indices may be produced with a small workload. The degree of quality fixation of the price index increases with the availability of attribute data. In addition, unlike the model pricing approach, this method does not require extensive expertise in building and civil engineering. However, increasing the number of attributes used for subdivision reduces the number of observations in the same stratum, often resulting in no price data in many strata (missing values), generating bias and noise.

(3) Hedonic method

---

[6] In this regard, the U.S. BLS(Bureau of Labor Statistics), together with experts (construction cost estimating companies), periodically reviews the building models to ensure that they are representative. Every month, when each contractor responds to the survey, they are asked to check in advance the factors affecting the determination of bid prices so that realistic prices are investigated. In addition, the collected prices are verified to ensure that they do not deviate from the actual prices, and cases in which the margin ratio fluctuates significantly due to changes in the content of the work (price changes due to quality changes) are discarded.

[7] For more information, see European Commission, Eurostat, OECD, and World Bank (2013).

This approach estimates the price of a product by considering it as an aggregate of the values of various performances and functions (a bundle of attributes) and using regression analysis. The commodity price is expressed as an equation consisting of a bundle of attributes, and this equation is called a "hedonic function." The regression analysis employs the collected price and attribute data, controlling for the effects of various attributes, and then creating a price index based on the estimated time-series dummies.[8]

Among the construction-related deflators currently used for GDP statistics overseas, only the U.S. uses the hedonic method. In the U.S., this approach is based on data from the Survey of Construction, a survey conducted by the Bureau of the Census on housing, including actual construction costs and data on the location, layout, and construction method of the housing sector.

In the hedonic method, it is easier to increase the degree of quality fixation compared to the stratification method because missing values do not occur. However, this approach requires knowledge of econometrics to estimate the hedonic function, and the estimation requires many attribute variables. When attributes may not be observable; therefore, misspecification problems may arise, such as endogeneity and omitted variable bias. Furthermore, it is necessary to periodically re-estimate the hedonic function, increasing the burden of producing price indices.

3. Estimation of Output-Type Construction Price Indices Using Big Data

3-1. Data

We construct an output-type construction price index for residential and non-residential buildings using the stratification and hedonic methods. Individual data from the Statistics of Construction Starts are used as the source material.

In Japan, the Building Standard Law requires the owner to submit a Notification of Construction Work to the prefectural governor when constructing a building. The MLIT compiles the Notification of Construction Work and publishes the Statistics on Building Starts. These statistics comprise all building investments. The survey data used in this

---

[8] For more information on estimating price indices using the hedonic method, see European Commission, Eurostat, OECD, and World Bank (2013).

study cover 7.92 million cases from January 2005 to December 2020 (an average of approximately 500,000 cases per year). A large amount of information on building prices and attributes may be obtained from the survey data. Table 1 shows the building attribute items listed in the survey forms (see Appendix 1 for details).

Table 1. List of Attribute Items in the Statistics of the Building Starts Questionnaire

number	Item Name	Sign and description
1	Year of survey	2005-2020
2	Survey month	01-12
3	Prefectural Number	01 to 47, Hokkaido to Okinawa
4	City, Town and Village Code	XXX
5	Intra-municipal serial number	XXXX
6	Scheduled construction period	01-99 (months)
7	the builder	1: country, 2: prefecture, 3: municipality, 4: company, 5: non-company organization, 6: individual
8	Structure	1: Wooden structure, 2: Steel-framed reinforced concrete structure, 3: Reinforced concrete structure, 4: Steel structure, 5: Concrete block structure, 6: Others
9	Building use	In addition to the classification based on the Standard Industrial Classification, classification is based on the use of offices and stores.
10	Construction type	1: New construction, 2: Extension, 3: Reconstruction
11	Capital stock classification	1: 10 million yen or less, 2: Over 10 million yen to 30 million yen or less, 3: Over 30 million yen to 100 million yen or less, 4: Over 100 million yen to 1 billion yen or less, 5: Over 1 billion yen (Only when the architect is "4: Company")
12	City planning classification	1：Urbanized area, 2：Urbanized control area, 3：Undefined urban planning area, 4：Quasi-urban planning area, 5：Under urban planning area and quasi-urban planning area
13	Building classification	1 to 9: Enter a series of numbers when there are two or more buildings in one construction report. The same number should be entered for the same building, and "9" should be entered for all buildings above "9"
14	Small number	If there are two or more houses in one building with different use relationships, enter the series number.
15	Number of floors above ground in new construction	01-99 (only when the construction type is "1: New Construction")
16	Number of basement floors in new construction	1 to 9 (only if the construction type is "1: New Construction")
17	Site area of new construction	m^2 (only if the construction type is "1: New Construction")
18	Total floor space	(As there is no obligation to report building work if the area is less than 10 m^2) it takes a value of 11 m^2 or more
19	Estimated construction cost	10,000 yen
20	Versatile or not	If 1, it indicates a multi-use building
21	Number of housing units to be removed	XXX
22	Relationship of use of retired housing	1: Owner-occupied house, 2: Rental house, 3: Salary housing
23	Building method	1: Conventional construction method, 2: Prefabricated construction method, 3: Frame wall construction method
24	Construction type	1: Newly established, 2: Other
25	Funds for new housing	1：Privately financed housing, 2：Public housing, 3：JHF housing, 4：Urban

		Renaissance Agency housing, 5：Others (only if the construction type is "1: New construction")
26	Type of housing	1: Dedicated housing, 2: Conjoined housing, 3: Other housing
27	Building (e.g., house)	1: Single-family houses, 2: Row houses, 3: Apartment houses
28	Usage restrictions, limitations	1: Owner-occupied house, 2: Rental house, 3: Salaried house, 4: Condominium house
29	Number of housing units	XXXX
30	Total floor area of the house	m^2

Note: Variables from 21 to 30 are only used for residential buildings.

Source: Compiled based on survey sheets from the Ministry of Land, Infrastructure, Transport, and Tourism's Compensation of Employees.

3-2. Overview of the Stratification Method

The stratification method involves the following four steps:

1. The data are divided into groups (called "stratification/subdivision") based on the attributes expected to affect the quality of a building, such as the structure of the building, construction method, and venue. Hence, the quality of the buildings within the same group is differentiated even though group members have similar characteristics.
2. The total amount of contract construction expenses, as well as the total floor area of a property, are calculated for the buildings included in the same group. Then, the unit price per floor area is obtained by dividing total expenses by the total floor area. This unit price represents the building price of the group.
3. The value of the unit price is transformed so that the index considers the average in FY 2011 as 100. The index is constructed for every group.
4. Finally, the indexed unit prices are aggregated by taking a weighted average based on the FY 2011 value of the total contract construction costs of each group.[9]

The index obtained after the above four steps becomes the final output building price index. In this study, we classify buildings into two categories: residential and non-residential buildings, which greatly differ in quality, and adopt quarterly aggregates.[10] In

[9] The construction price index is a fixed-standard Laspeyres index. As described later, the output-type indices estimated in this study are compared with the input-cost indices, the Ministry of Land, Infrastructure, Transport, and Tourism's "Construction Cost Deflator;" as the figures for the FY 2015 base start in the second quarter of 2011, the figures for the FY 2011 base are used.

[10] The available attribute items are different for residential and non-residential properties, and the percentage of missing values can be greatly improved by using quarterly aggregation instead of monthly aggregation. For details, please contact the authors.

addition, for missing price data, we examine the following five typically used imputation methods: (1) the last observed price data, (2) the price data of the same period in the previous year, (3) the average price of non-missing data, (4) the average quarter-on-quarter growth rate of non-missing data, and (5) the average year-on-year growth rate of non-missing data.[11,12] We input the missing price data using the last observed price data because this approach results in the smallest fluctuations in the estimated price index and better captures essential turning points for building prices.

For stratification, items that greatly affect the quality of the building and have an impact on the unit price of the building should be selected. Specifically, for residential buildings, the "construction method" (prefabricated and two-by-four, among others), "construction method" (single-family house and apartment house, among others), "structure" (wooden and reinforced concrete, among others), "prefecture (region)," and "use relationship" (owner-occupied and rental house, among others) are selected. In the case of non-residential buildings, "building use" (e.g., industry of the company that built the building), "structure," and "prefecture (region)" are used (Table 2).

Table 2. Attribute Items in the Stratification Method

	attribute (specification) item	contents
Housing	Building method	Prefabricated, two-by-four, other (conventional construction method)
	Building (e.g., house)	Single-family houses, row houses, and apartment buildings
	Structure	Wooden construction, steel-framed reinforced concrete construction, reinforced concrete construction, steel construction, and concrete block construction, among others
	Administrative divisions of Japan	47 Prefectures
	Usage restrictions, limitations	Owner-occupied houses, rental houses, salaried housing, and condominiums
Non-housing	Versatile or not	Multi-purpose buildings, among others
	Purpose	Standard Industrial Classification (Middle Classification)

[11] The results are omitted here for the sake of brevity. For details, please contact the authors.

[12] Regarding "private rents" in the Consumer Price Index, the index is created by supplementing the rents for the most recent month horizontally for rented houses that became vacant due to renters moving out, thus becoming missing values.

structure	Wooden construction, steel-framed reinforced concrete construction, reinforced concrete construction, steel construction, and concrete block construction, among others
Administrative divisions of Japan	47 Prefectures

3-3. Overview of the Hedonic Method

The dependent variable of the hedonic function is the unit price per floor area of the estimated construction expenses (estimated construction expenses/floor area). In this study, we use the logarithmically transformed value, in line with Diewert (2003). We adopt a rolling estimation method, in which the estimation is performed while shifting the window up to 12 months. This is the recommended method for estimating housing price indices by the European Commission, Eurostat, OECD, and the World Bank (2013) and has the advantage of reflecting time-series changes in parameters.[13] All available attribute items are used as explanatory variables, and a one-sided log-linear hedonic function is employed, as follows:[14]

$$\log p_i = \alpha + \sum_{j=1}^{n} \beta_j x_{i,j} + \sum_{k=1}^{m} \delta_k d_{i,k} + \sum_{t=1}^{T} \gamma_t TD_{i,t} + u_i, \quad \ldots \qquad (1)$$

p_i: unit price per square meter of building i;
α: constant term;
$x_{i,j}$: j^{th} attribute of building i (numerical value);
β_j: parameter of numerical data;
$d_{i,k}$: kth attribute of building i (dummy);
δ_k: parameter of the dummy variable;
$TD_{i,t}$: survey month t of building i (time dummy);
γ_t: time dummy parameters (representing quality-adjusted prices);
u_i: error term.

It is difficult to assume a linear relationship between the qualities of buildings due to the wide distribution of the number of stories. Therefore, in addition to setting multiple

[13] The rolling window hedonic method was developed in 1997 as an estimation method for the MSCI-IPD Residential Property Price Index. The details were later published in Shimizu, Takatsuji, Ono, and Nishimura (2010). The method is also used in the "Public Property Price Index" published by the Ministry of Land, Infrastructure, Transport and Tourism.

[14] The estimation period ranges from January 2005 to December 2020, 181 times the rolling estimation in a 12-month window frame; no explanatory variable’s coefficient is found to be not significant throughout the whole period.

dummy variables in a nonparametric manner, a piecewise linear function is also introduced.[15] This function divides the number of floors into multiple categories and assumes linear relationships within each category. For example, in the case in which the number of floors is divided into three categories, we obtain:[16]

$$f_S(S_i) \equiv D_{S,i1}\lambda_1 S_i + D_{S,i2}[\lambda_1 S_1 + \lambda_2(S_i - S_1)] + D_{S,i3}[\lambda_1 S_1 + \lambda_2(S_2 - S_1) + \lambda_3(S_i - S_2)], \quad (2)$$

where S_i is the building i' s number of stories above ground (S_1, S_2 are the maximum number of stories in Category 1 and Category 2, respectively), $D_{S,i}$ is a dummy variable that takes a value of 1 if the number of stories above ground falls into each category, and λ is the coefficient through which each story category affects the unit price per gross floor area.

In cases in which a dummy variable is not continuously observed throughout the rolling window, an adjustment is made to redefine it into a "wider range of dummy variables" integrated with other items. This adjustment is made since the inclusion or exclusion of these explanatory variables (depending on the point in time of the estimation) may lead to significant changes in the estimated values of the coefficients.[17]

4. Estimation Results of the Output-Type Building Price Index

4-1. Output-type Building Price Index and Construction Cost Deflator

This section compares the output-type building price indices estimated by the stratification and hedonic methods with the construction cost deflator, an input-cost index. We evaluate the estimation results focusing on six categories, namely wood-frame, reinforced concrete (RC), and steel-frame (S) construction for residential and non-residential buildings, respectively, which account for a large share of the total building

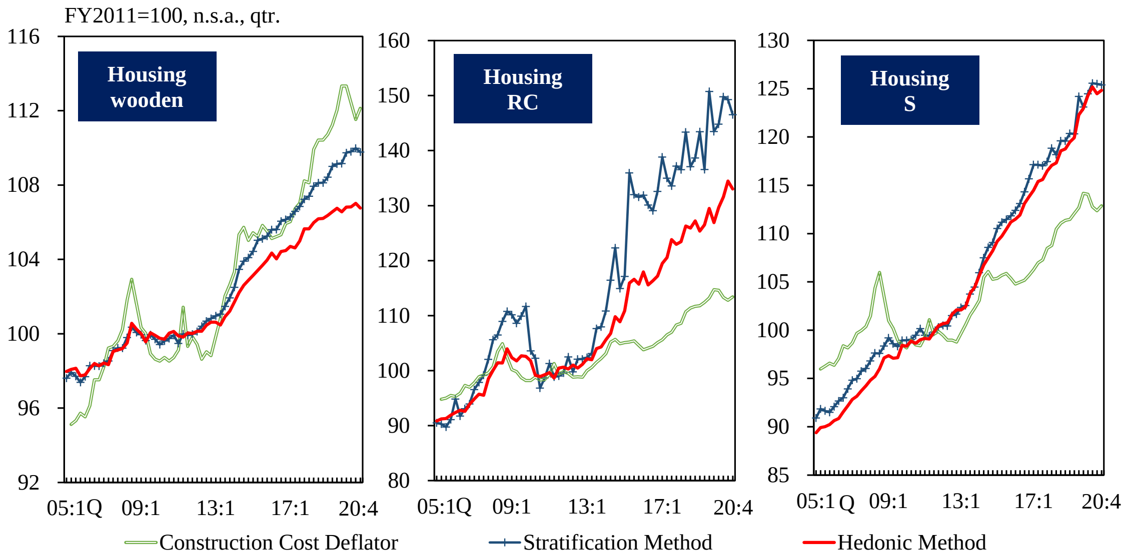
[15] For details on the application of piecewise linear function, please refer to Shimizu et al. (2014), Diewert and Shimizu (2015, 2016, 2017), and Shimizu and Diewert (2018).

[16] The use of piecewise linear functions allows us to consider cases in which the impact of increasing the number of floors from a one-story building to a two-story building is different from the impact of increasing the number of floors from a 20-story building to a 21-story building.

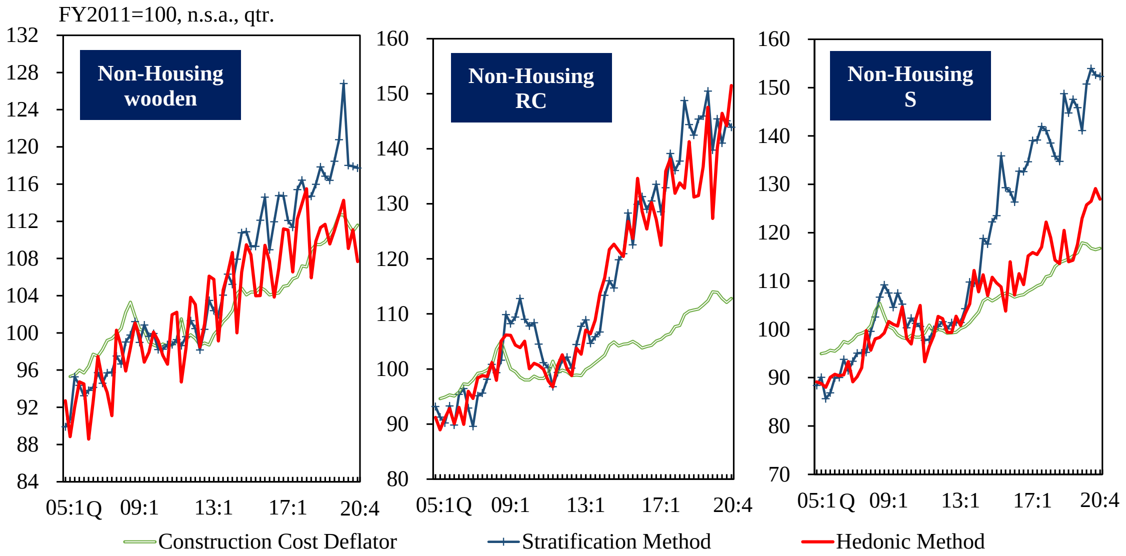
[17] Regarding the dummy variables, adjustments are made, for example, to recombine the dummy variable representing Aomori Prefecture into a regional dummy for the Tohoku region, and to redefine the usage dummies by grouping them by industry instead of dividing them by use, such as warehouses and stores.

stock.

The estimation period ranges from the first quarter of 2005 to the fourth quarter of 2020. It includes an expansionary phase, approximately until 2007, a sharp recessionary phase triggered by the Lehman shock in 2008–2009, and a long expansionary phase from the fall of 2012 to the fall of 2018. From autumn 2012 onward, construction investment has raised substantially, and the construction sector's profits have increased markedly due to an improvement in the order environment.[18] In terms of prices, the sharp rise in resource prices up to approximately 2008 has increased the prices of construction materials, followed by a decline in material prices in 2009, and a gradual rise in prices since 2013. These phenomena reflect substantial price changes, such as increases in labor costs due to the persistent labor shortage in the construction sector and rises in the profits of construction companies. Hence, we evaluate changes in the construction price index that reflect such variations in the economic conditions. The estimated indices are compared with the construction cost deflator (Figure 1).



[18] In addition to these factors, the recession caused by the spread of COVID-19 and the subsequent rise in housing demand have caused lumber prices to start rising from approximately July 2020, leading to a "wood shock." The Bank of Japan's "Corporate Goods Price Index" confirms that the impact of rising import prices for materials has become more significant since the beginning of 2021, but the impact has not been as pronounced during the analysis period (until the end of 2020).



Source: Ministry of Land, Infrastructure, Transport, and Tourism; construction cost deflator; calculations by the authors.

Figure 1. Output-Type Building Price Index

The output-type construction price indices obtained by the stratification and hedonic methods show a larger increase than the input-cost construction cost deflator for all series except for housing and wood construction (housing and RC construction, housing and S construction, non-housing and wood construction, non-housing and RC construction, and non-housing and S construction). In many cases, the gap between the construction cost deflator and the output-type construction price index was almost negligible until 2012, but from 2013 onward, the gap has gradually increased. In the economic expansion phase, since 2013, the true construction prices captured by the output-type construction price index are higher than those captured by the existing input-cost price indices (the construction cost deflator and SNA's index). The results suggest that the estimated output-type construction price index may effectively overcome the measurement difficulties outlined in this study.

In the late 2000s, the timing of the increase in the indices based on both the stratification and hedonic methods, especially for residential and RC construction, lags behind the increase in the construction cost deflator, and the fluctuation is modest. This result indicates that construction companies cannot immediately pass on changes in materials and labor costs reflected in the construction cost deflator to building owners in a competitive environment; hence, they negotiate prices over time and reflect them in their contract prices.

Finally, comparing the indices obtained by the stratification method with those generated by the hedonic method, we find that in some cases, they show approximately the same increase (residential/S and non-residential/RC); however, a certain gap is observed between the price index obtained using the stratification method (residential/wooden) and that obtained by the hedonic method. The former increases more than the latter, and the gap between the two is, at times, large (residential/RC, non-residential/wooden, and non-residential/S) (Table 3).

Table 3. Discrepancies between the Stratification Method, Hedonic Method, and Construction Cost Deflator–Housing/Non-Housing and Structure

Housing/ Non-housing	Structure	Stratification v.s. Construction Cost	Hedonic v.s. Construction Cost	Stratification v.s. Hedonic
Housing	wooden	medium gap	medium gap	medium gap
	reinforced concrete construction	**large gap**	**large gap**	**large gap**
	steel construction	medium gap	medium gap	small gap
Non-housing	wooden	**large gap**	medium gap	**large gap**
	reinforced concrete construction	**large gap**	**large gap**	small gap
	steel construction	**large gap**	medium gap	**large gap**

As described above, among the six series of output-type indices, the gap between the two indices is large for three series: residential/RC construction, non-residential/wood construction, and non-residential/S construction. This discrepancy may be due to inappropriate selection of attributes for stratification and subdivision in the stratification method and insufficient quality adjustment, among others. In addition, the hedonic method may suffer from estimation bias. In the next subsection, we further discuss this point.

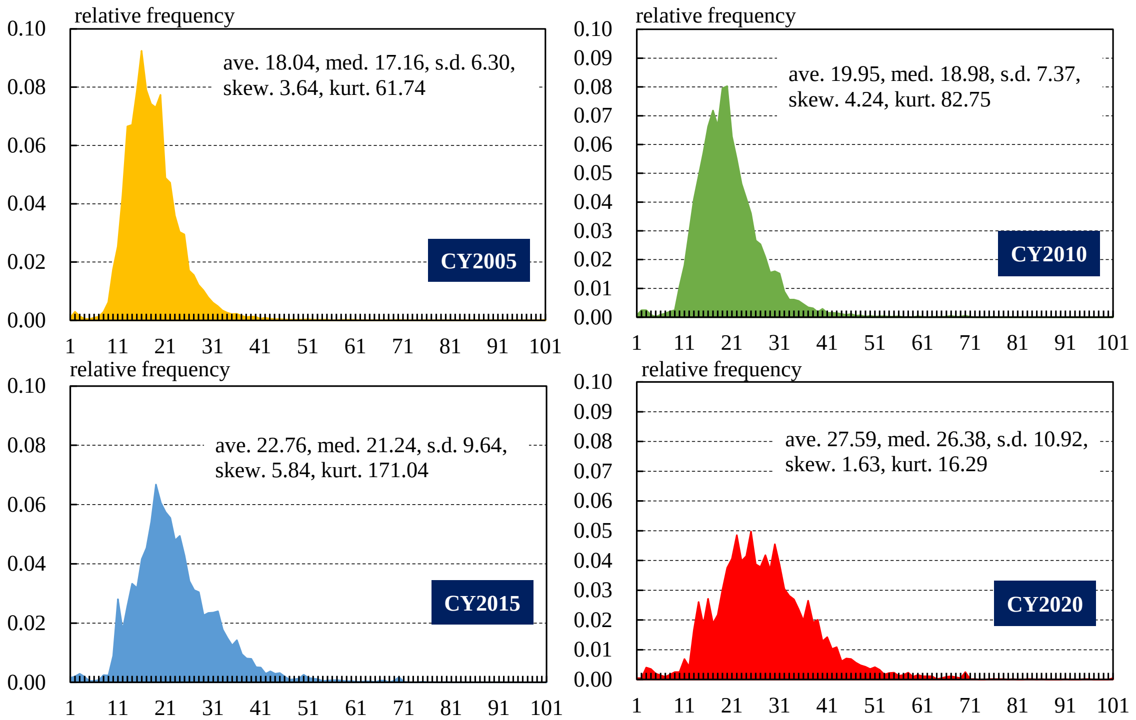
4-2. Evaluation of Output-Type Building Price Index: Dispersion Factors

4-2-1. Increasing variability in residential/RC and non-residential/S constructions

Among the three series in which substantial gaps are observed, for residential/RC and non-residential/S constructions, the unit price per total floor area of the estimated construction expenses has been increasing in recent years (Figure 2).[19]

This result indicates that the variation in the quality of buildings has been large in residential/RC and non-residential/S. This phenomenon may be due to an increase in the number of high-rise condominiums such as tower apartments and the larger variation in buildings constructed with steel frames owing to the recent improvement in construction technology.

To adjust for the effects of these increased variations, in the stratification method, the elements detailed in Table 4 are added as items for subdivision and stratification, and the stratification items are further subdivided to improve the degree of quality fixation of attributes.



Note: The unit of x-axis is ten thousand yen except the last bin which shows over one million.
(Source: Ministry of Land, Infrastructure, Transport and Tourism and calculations by the authors)

Figure 2. Distribution of Unit Price per Square Meter in Residential RC Construction

[19] Although only the graph for residential/RC is shown, the same trend is confirmed for non-residential/S.

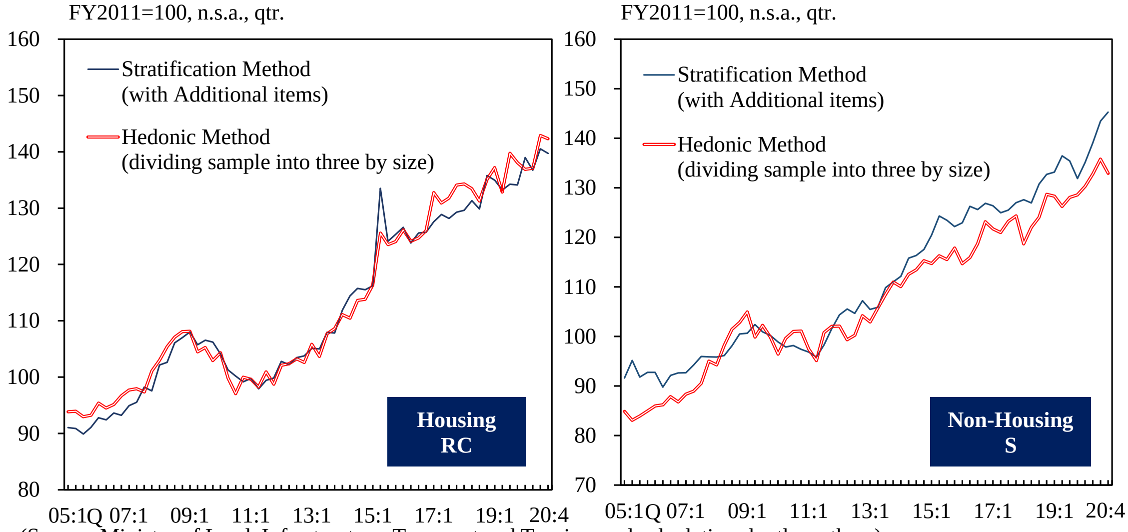
Table 4. Additional Items in the Stratification Method

Housing/ Non-housing	Baseline	Addition of stratification and subdivision items
Housing	Building method, construction method, structure, prefecture, and use relationship	Baseline; Builder, capitalization category, city planning category, basement, retired housing, funds for new housing, use (building with residential industrial use)
Non-housing	Multi-use or not, use, structure, and prefecture	Baseline; Architect, capitalization category, city planning category, underground

In addition, in the hedonic method, we divide the data by setting a total floor area as a threshold to estimate the function for each data point.[20] By dividing the sample by the size of the buildings, differences in the coefficients of the explanatory variables in the hedonic estimation may be considered, preventing small buildings, which have a small share in the value of the estimated construction expenses but a high share in the sample size, from having excessive influence.

Figure 3 compares the indices obtained by the stratification method with the addition of subdivision and stratification items and the indices obtained by the hedonic method with sample division by total floor area. Comparing Figure 3 with Figure 1, we show that the difference between stratification and hedonic indices is much smaller than that obtained before dividing the sample for both residential/RC and non-residential/S.

[20] Referring to the distribution of the number of observations, we estimate the cases divided into three categories for residential and RC structures: total floor area of 200 m^2 or less, over 200 m^2 to 800 m^2, and over 800 m^2; for non-residential and S structures: total floor area of 100 m^2 or less, over 100 m^2 to 300 m^2, and over 800 m^2.



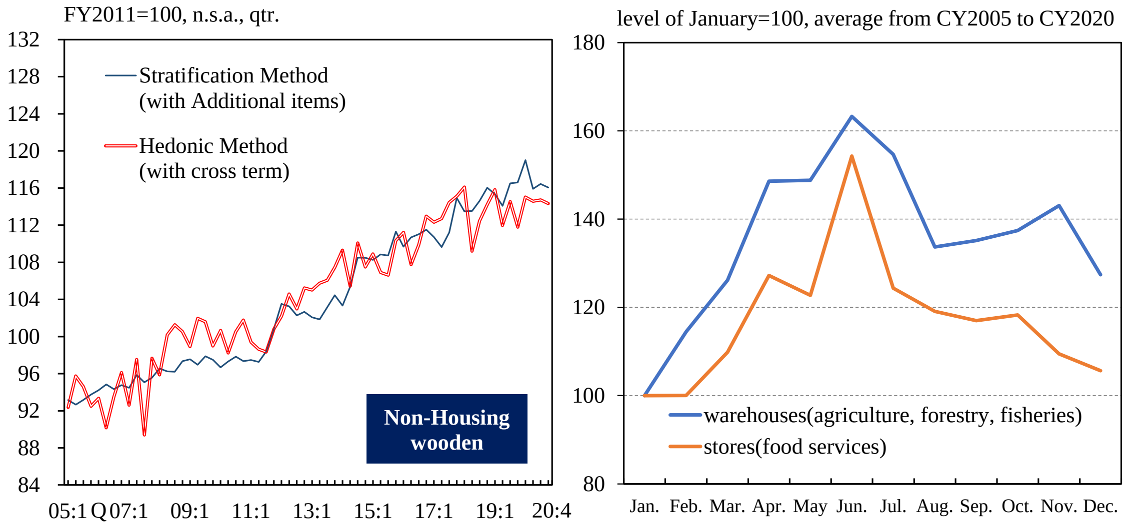
(Source: Ministry of Land, Infrastructure, Transport and Tourism and calculations by the authors)

Figure 3. Stratification, Hedonic method: Additional Estimation Index 1

The degree of quality fixation in the stratification method is improved by increasing the number of items to be subdivided and stratified, enhancing the accuracy of the index. In addition, by dividing the sample addressed by the hedonic method, the price trends of buildings with a high rate of price increase and a large scale, such as tower condominiums, are reflected by a more appropriate weight, narrowing the gap between the two methods.

4-2-2 Effect of seasonality on non-residential and wooden structures

Among the three series in which index gaps are observed, the non-housing and wood-frame series show large quarterly fluctuations in both indices, suggesting that seasonality is likely observed (Figure 1). By addressing the number of observations for each month for warehouses in agriculture, forestry and fisheries industry (corresponds to 314 classification number of types of dwellings, industries, and use) and stores in food services industry (classification number: 532), which have a large share in the non-residential/wooden category (Figure 4, right), we observe seasonal fluctuations with a peak in June. The seasonal changes in the composition of non-residential/wooden buildings may affect the estimated indices because the unit price per square meter for seasonally built constructions such as beachside houses and small cottages used for storing harvests is significantly lower than for other buildings.



(Source: Ministry of Land, Infrastructure, Transport and Tourism and calculations by the authors)

Figure 4. Left: Stratification, Hedonic Method; Additional Estimation Index 2
Right: Number of Observations for Use/Application Categories
Applicable to Non-Housing/Wooden

Based on these characteristics, in the stratification method, the index is created by adding stratification items in the same way as in the two series described in the previous subsection. In the hedonic function, "use of buildings" dummy variables, such as "office" and "store", etc., are added in addition to the industry of the building owner company, and the interaction term between the industrial/usage dummy and the time dummy is employed to capture seasonal changes in the composition of buildings within each industry and use category (Figure 4, left). The results show that the gap between the index obtained by the stratification method with the addition of stratification items and the hedonic index with the addition of the cross term is reduced. In addition to the increase in the degree of quality adjustment in the stratification method, the effect of seasonal variation is removed by adding the cross term in the hedonic method; thus, the fluctuation in the index is smaller than that shown in Figure 1. The index is thus considered more appropriate.

5. Conclusion

This study builds an output-type building price index using survey data from the Statistics of Building Starts, using the stratification and hedonic methods.

By considering changes in the six series of indices for residential and non-residential buildings categorized by the structure type, we obtain an output-type building price index with a reasonable level of accuracy. Except for housing/wooden, during the sample period, all series (housing/RC, housing/S, non-housing/wooden, non-housing/RC, and non-housing/S) have increased at a higher rate than the construction cost deflator, the input-cost price index, and currently official statistics. This trend has been particularly pronounced recently, especially since 2013. The output-type building price indices obtained by the stratification and hedonic methods reflect that profits in the construction sector have been increasing due to the recent improvements in the environment for construction orders. As such, the output-type building price index provides valuable information.

We confirm that the current input-cost price indices may not be sufficient to improve the accuracy of real construction investment in Japan and accurately reflect the activities of the construction sector. A new, highly accurate output-type construction price index may better serve this scope.

The stratification method allows a higher degree of quality fixation. In contrast, the hedonic method requires the accumulation of know-how and the repeated estimation every month and every quarter. In addition, the stratification method requires relatively small compiling costs, as it may be obtained by stratifying and subdividing information based on major attribute items, calculating the unit price per gross floor area for each stratum, and taking the weighted average.

This study's results indicate that the indices obtained by the stratification and hedonic methods for the three series of residential/wooden, residential/S, and non-residential/RC show approximately the same level of increase. However, for the three series of residential/RC construction, non-residential/wooden construction, and non-residential/S construction, a certain discrepancy is observed between the two indices. These gaps may be reduced by refining the estimation. The stratification method may be used to create an output-type building price index at a low cost, providing support for the practical application of such an index in the future.

Despite its methodological contribution, the current study has some limitations.

First, the proposed price index does not reflect improvements in the quality of labor, which are expected to have a significant impact in the medium to long term. The attribute items of the Statistics of Construction Starts used do not reflect improvements in building quality associated with increased quality of labor due to data limitations. Second, the estimated construction expenses of a building provided by the Statistics of Construction Starts are estimated at the time when construction began and not at the time of completion; hence, they do not reflect design changes after the construction start. The method for converting the price index into an accrual-based index to match the SNA standards for estimating construction investment is also worth further investigation. In addition to further verification of the estimation method, the above issues should be examined in the future.

References:

Cabinet Office, Government of Japan, *Explanatory Notes on the Methodology for Estimating National Accounts (Annual Estimates), 2015 Standard Version*, November 27, 2020; revised January 20, 2021. Available at: https://www.esri.cao.go.jp/jp/sna/data/reference1/h27benchmark/pdf/kaisetsu_20210120.pdf

Diewert, W. E., "Hedonic Regressions: A Review of Some Unresolved Issues," Paper presented at the 7th Meeting of the Ottawa Group, Paris, May 27–29, 2003.

Diewert, W. E. and C. Shimizu, "Residential Property Price Indices for Tokyo," *Macroeconomic Dynamics,* 19(8), 1659–1714, 2015.

Diewert, W. E. and C. Shimizu, "Hedonic Regression Models for Tokyo Condominium Sales," *Regional Science and Urban Regional Science and Urban Economics*, 60, 300–315, 2016.

Diewert, W. E. and C. Shimizu, "Alternative Land Price Indices for Commercial Properties in Tokyo," Discussion Paper 17–07, Vancouver School of Economics, University of British Columbia, 2017.

Diewert, E and C. Shimizu (2021), “Residential Property Price Indexes: Spatial Coordinates versus Neighbourhood Dummy Variables,” *Review of Income and Wealth*, forthcoming, DOI:10.1111/roiw.12534.

European Commission, Eurostat, OECD, and World Bank, *Handbook on Residential Property Price indices (RPPIs)*, Publications Office of the European Union, Luxembourg, 2013. Available at: doi:10.2785/34007

Fukao, K., T. Kameda, K. Nakamura, R. Namba, and M. Sato, "Measurement of Deflators and Real Value Added in the Service Sector," Economic Analysis, No. 194, Economic and Social Research Institute, Cabinet Office, Government of Japan, 2017.

Ministry of Land, Infrastructure, Transport, and Tourism, *Outline and Revision of the Construction Cost Deflator (FY2015 Base)*, Construction Economics and Statistics Office, Policy Bureau, Ministry of Land, Infrastructure, Transport and

Tourism, 2021. Available at: https://www.mlit.go.jp/sogoseisaku/jouhouka/content/001411495.pdf

OECD, *Construction Price Indices: Sources and Methods*, OECD, Eurostat, 1997.

Shimizu, C., and W. E. Diewert, "Alternative Land Price Indices for Commercial Properties in Tokyo -Comparison with Different Data Sources," Paper presented at the 5th Annual Conference of the Society for Economic Measurement, Xiamen University, China, June 9, 2018.

Shimizu,C., K. Karato, and K. G. Nishimura, "Nonlinearity of housing price structure— Assessment of three approaches to nonlinearity in the previously owned condominium market of Tokyo," *International Journal of Housing Market and Analysis*, 7(4), 459–488, 2014.

Shimizu, C., H. Takatsuji, H. Ono, and K. G. Nishimura, "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices," *International Journal of Housing Markets and Analysis*, 3(4), 2010.

Appendix 1

Scope and Details of the Data in the "Compensation of Employees"

Data Scope

The Statistics on Construction Starts are based on the Notification of Construction Work" submitted when constructing a building, such as houses and non-residential buildings.

Appendix Table 1. Left: Methodology for Developing Valid Price Indices; Right: Construction Investment in 2019

Target Property	Private sector	Public
Housing	Stratification method/ hedonic method	
Non-residential・Buildings		
Construction・Repair	Model pricing approach (survey of construction companies)	Model pricing approach (e.g., survey of construction companies or use of public data)
Civil Engineering		

Target property	Private sector	Public	Total
Housing	19.3 trillion yen (28%)	0.3 trillion yen (0%)	19.6 trillion yen (28%)
Non-residential・Buildings	11.8 trillion yen (17%)	2.8 trillion yen (4%)	14.6 trillion yen (21%)
Construction・Repair	10.1 trillion yen (15%)	1.8 trillion yen (3%)	11.8 trillion yen (17%)
Civil Engineering	6.7 trillion yen (10%)	16.3 trillion yen (24%)	23.0 trillion yen (33%)
Total	47.9 trillion yen (69%)	21.2 trillion yen (31%)	69.0 trillion yen (100%)

Note: The amount of construction investment (69.0 trillion yen in 2019) is based on fixed capital formation (for construction investment) in the SNA (65.0 trillion yen) plus "maintenance and repair work" (4.0 trillion yen), treated as intermediate consumption in the SNA. Construction and repair include "renovation and repair work" as well as "maintenance and repair work."

Source: Authors' estimate based on the National Accounts published by the Cabinet Office and the Survey of Construction Work Statistics published by the MLIT.

Of the total construction investment (69.0 trillion yen) in 2019, the amount of investment in residential and non-residential buildings covered by the "Compensation of Employees" is 34.2 trillion yen, 50% of the total (See Appendix Table1).[21]

Data Details

[21] In contrast, it is difficult to collect a large amount of data on prices and attributes for investment in civil engineering structures and building repair work, which are not covered by the Statistics on Building Starts, and it is difficult to use the stratification approach and the hedonic method. For civil engineering and building repair work, it is necessary to use the model price approach.

Table 1 shows the attributes surveyed in the Compensation of Employees. Items 1–20 are surveyed for all buildings, and items 21–30 are also surveyed for buildings that fall under the housing category.

In this study, we estimate the price index based on the unit price per square meter, calculated as the estimated construction expenses divided by the total floor area. The estimated construction expenses under the Compensation of Employees are defined as the sum of the construction expenses of the main body of the building and the construction expenses of the building facilities, which corresponds to the contract amount of the construction contract that the owner of the building orders from the construction company. Therefore, it is possible to obtain the market transaction price (output price) of buildings, a base that includes the construction company's profit in addition to the cost of materials and labor required for construction, making the data suitable for the proposed analysis.[22] As a large number of attributes, such as the structure, number of floors, use, and region of the building are also surveyed, and the dataset is large, as it is a complete survey, this information may be used to adjust the quality of the building in the stratification and hedonic methods.

In this study, we do not use data on additions and renovations, but we limit the analysis to new buildings because attributes such as the number of stories above ground and site area, expected to have a large impact on the quality of buildings, are surveyed only for new buildings. In addition, as the Compensation of Employees is a complete source of statistics, we do not clean the data in principle. However, as individual data may have been incorrectly filled, we discard such information when analyzing the price indices.[23] The total number of samples from January 2005 to December 2020 is 7,924,280; of them, 7,149,919 are residential and 783,361 are non-residential.

[22] It should be noted that the estimated construction expenses in the Statistics on Construction Starts is the value at the time the notification of construction work is submitted, not the actual amount of construction expenses at the time of completion, nor is it converted to a progress basis.

[23] Adjustments to the individual data include: 1) exclusion of data with inconsistent information, such as data that correspond to housing but do not have data points for housing survey items, 2) exclusion of data with extremely large values for the number of stories above ground and building-to-land ratio, and 3) corrections such as connecting city codes and industry classifications.

Appendix 2

Estimation Results for the Hedonic Method

Based on Equation 1, the following adjustments are made in the estimation of the hedonic function based on the characteristics of the data.

First, given the large variation in prices and number of observations by structure, we divide the data into four residential categories–residential/wooden, residential/reinforced concrete (RC), residential/steel-frame (S), residential/other, and four non-residential categories–non-residential/wooden, non-residential/reinforced concrete (RC), non-residential/steel-frame (S), and non-residential/other. Namely, we estimate the hedonic functions after dividing the data into those eight categories.

Second, we use three explanatory variables for the numerical data–the estimated construction period, site area, and total floor area for both residential and non-residential buildings, and two additional variables for residential buildings–the number of housing units and total disposed housing dummy × floor area. As both coefficients are significant, they significantly contribute to the quality adjustment. Regarding the number of basement floors, we use two types of dummies–a dummy for buildings with basements and a dummy for buildings with two or more floors among them, as the number of buildings with three or fewer basement floors was approximately 400 units.

The results of a total of 181 rolling estimations from January 2005 to December 2020 show that the coefficients on most explanatory variables are significant, although significance changes depending on the estimation period. Although the number of explanatory variables is large, the problem of degrees of freedom is not observed thanks to the large size of the Building Starts Statistics, an all-year-round survey.

In terms of adjusted R^2, the minimum is 0.08, and the maximum is 0.51, suggesting a certain limit to the explanatory power of the hedonic function, in addition to the influence of the large sample size. This phenomenon may be due to attributes that are not covered by the Building Starts Statistics but affect the quality of buildings (e.g., the seismic strength of the building, the grade of the interior, attached equipment, and building materials, among others). As buildings and real estate are highly individualized, as suggested by existing studies (Diewert and Shimizu (2015), (2016), (2017), (2021)), the commonly used explanatory variables alone may not have sufficient explanatory power.

To verify the robustness of the hedonic function, we check four points: 1) outliers in the questionnaire data of the Building Starts Statistics, 2) changes in the width of the window (12 months) adopted in the rolling estimation, 3) multicollinearity of explanatory variables, and 4) nonlinearity of explanatory variables. Although the results are omitted, the effects of all four of the above points on the estimation results of the hedonic function and the building price index are minor.[24]

[24] For more information, please contact the authors.