CSIS Discussion Paper No.93

Estimation of Hedonic Single-Family House Price Function Considering Neighborhood Effect Variables[*]

Chihiro Shimizu[†]
Reitaku University

January 13, 2009

-Abstract-

In this study, hedonic house price functions are formulated using the single-family house market in the 23 wards of Tokyo as the subject. In the formulation of hedonic models, the generation of omitted variable bias is thought to occur in cases when, in addition to locational factors (i.e., factors involved in decision making when buying a property) and building structures which affect the house prices, local environmental variables and the individual characteristics of house buyers, such as their income, are not taken into consideration. However, since it is difficult to obtain local environmental information in a small neighborhood unit and to observe individual characteristics of house buyers in the property market, these variables have not been sufficiently considered in previous studies. In the current study, I aim to improve the hedonic model by incorporating local environmental factors and data on family income in a small neighborhood unit using a geographic information system (GIS). I demonstrate that, without considering these variables, non-negligible levels of omitted variable bias are generated in the variables that are major factors in determining house prices, such as ground area, front road width, distance to the nearest station, and the travel time to the central business district.

Keywords: Hedonic model, identification, omitted variable bias, Neighborhood Effect, Geographic Information System

JEL Classification: C31 - Cross-Sectional Models; Spatial, R31 - Housing Supply and Markets.

---

[*] I would like to thank Kiyohiko G. Nishimura(Bank of Japan), Atsuyuki Okabe,Yasushi Asami,Takatoshi Tabuchi(University of Tokyo), Koji Karato(Toyama University) and YongHeng Deng(University of Sothern California) for discussions and comments.This research is partly supported by the Grant in Aid for Research No. 19653025(Representative is Yasushi Asami) by Ministry of Education, Science and Culture in Japan and CSIS(Center for Spatial Information Science of University of Tokyo) joint research.

[†] Associate Professor, Department of International Economics and Business, Reitaku University, Visiting Research Fellow, Center for Spatial Information Science (CSIS), University of Tokyo
Tel: +81-(0)4-7173-3439(Laboratory), E-mail: cshimizu@reitaku-u.ac.jp

1. Objective of the study

Economic growth and the progress of urbanization have promoted high-density land use and the construction of high-rise buildings, thereby generating urban problems such as traffic jams, the obstruction of insolation and ventilation, and the impairment of view. To cope with these problems, policies for the preservation of high-quality urban space through the regulation of land use and other measures have been actively adopted within cities. However, policies such as land-use regulations cannot always be achieved with the agreement of citizens because uncertainties arise about the effects of such policies. Accordingly, various attempts to measure the effect of a policy such as land-use regulation in terms of economic values have been made. The most representative technique for such attempts is a hedonic model that focuses on property values.

However, in the formation of hedonic models, as indicated by Ekeland et al. (2004), several problems still remain. The most important problem is the identification of estimated variables. Namely, in the estimation using a hedonic model, price is determined as a result of the locational activity (decision-making of a family or a company to buy a property or land) of a standard household (family income) (Rosen (1974)); whereas in an actual city, heterogeneity of family income should be taken as a prerequisite (Kanemoto and Nakamura (1986)). In this case, data specific to the individual such as income, family types and so on is required. In addition, when a hedonic model is constructed for a wider space, neighborhood variables that can explicitly handle the differences among neighborhoods must be incorporated as explanatory variables. However, in the actual construction of hedonic models, very few studies have appeared in which these variables are incorporated explicitly (Asami and Gao (2006)). This lack is due to the problem of bias as a result of omitted variables in the estimation, which occurs because of the impossibility of incorporating all variables that must be considered (Ekeland et al. (2004)).

However, if information on the attributes of a family such as income cannot be obtained directly, then it may be natural to attempt to replace the information with collected statistical values and other observable variables. In addition, with the provision of spatial information and progress in spatial information processing technologies, it has become possible to observe data in a small neighborhood unit which represent regional characteristics at a level much higher than before.

Under such circumstances, in this study, starting from a hedonic model in which only the building attributes and locational characteristics are considered, I attempt to construct a model wherein regional characteristics in a small neighborhood unit and family income attributes, which have not been considered before(in Japan), are taken into account. Since such regional characteristic variables and the family income attribute variables affect each other, they show externalities. Therefore, I call them neighborhood effect variables. By comparing a model incorporating these variables with a model that does not include them, I clarify whether or not omitted variable bias is generated.

Results showed that, in the model considering neighborhood effect variables, the explanatory power was improved compared with the simple linear model. At the same time, each of the neighborhood effect variables, which served as subjects of urban policies, was adopted with statistical significance. As neighborhood effect variables, I adopted the conditions of the use of land and buildings and the building density in a unit of 500 x 500 (m^2) mesh, as well as average gross floor area and the ratio of office-worker households, which are both proxy indexes representing income level and are collected in the national census. In addition, I measured the road traffic noise at each point of the unit.

Furthermore, to consider regional characteristics that cannot be explained by the incorporation of these variables, I incorporated latitude and longitude coordinates as variables, and expanded the original hedonic model. In addition, I further improved the model by assuming a close relationship between locational characteristics and land-use density. Namely, I obtained the land-use density in each zone defined by urban planning, and incorporated the cross terms of the land-use density with the ground area, floor space, front road width, age of the building, time to the nearest station, travel time to the central business district (CBD), and the floor area ratio, all of which have significant affect on house price, in the model to obtain an approximate model that controls fixed effects.

As a result, I found that non-negligible bias was generated in the estimated parameters of not only the variables representing locational characteristics such as the front road width, time to the nearest station, the a dummy variable for bus transportation (bus dummy), and travel time to the CBD, but also the age of the building.

2. Hedonic model with consideration given to neighborhood externalities

2.1. Effect of neighborhood externality on price

Locational activities are affected by environmental factors in the broad sense of the term, such as accessibility to transportation, possibility of development, or the conditions imposed on the use of buildings in the neighborhood. In addition, the possible use of buildings is restricted by land-use regulations. One purpose of land-use regulations is to ensure high-quality urban space by separating the land in accordance with its different uses and by eliminating effects that induce external diseconomy. If these factors have some effect on people's locational activities, these factors are also reflected in house prices. Namely, "capitalization" through so-called "voting with your feet" is generated(Tiebout (1956)).

Thus, the house price is determined by factors such as the living environment resulting from land-use regulations, locational characteristics such as travel time to the CBD, the conditions imposed on local land use, the condition of the building including its age after construction and floor space. Various differentiations can be made in terms of performance and function. Additionally, as formulated by Rosen (1974), a bid-rent function, which constitutes a hedonic model, is determined by the characteristics of family income.

However, in many cases of hedonic-model estimation, characteristics of family income are ignored, and estimation is often made using only variables that can be obtained. Additionally, differences in geographical attributes are in most cases replaced by the travel time to the CBD or by dummy variables representing the local environment. However, when a gap appears in the price structure in terms of space and unobservable geographical differences are a factor, the use of these proxy variables is not sufficient. In the parametric polynominal expansion model proposed by Jackson (1979) or in geographically weighted regression (for example, Zuang (2006)), the formulation of functions with consideration given to spatial heterogeneity is possible; however, back ground causes of heterogeneity in the conventional hedonic model cannot be dealt with explicitly.

Under such circumstances, in this study, I aim to improve the hedonic model by organizing attribute information on local land-use conditions and regional household characteristics as neighborhood effect factors.

2.2. Estimation model

In this study, I set up the following hedonic model using the market of single-family houses and land in the 23 wards of Tokyo as the analytical subject.

$$\log DP/GA = a_0 + \sum_h a_{1h} X_h + \sum_i a_{2i} Z_i + \sum_j a_{3j} \cdot LD_j + \sum_k a_{4k} \cdot RD_k$$
$$+ \sum_l a_{5l} \cdot TD_l + \sum_m a_{6m} \log V_m + \sum_{h,m} a_{7h,m} X_h \cdot V_m + a_8 u + a_9 v + \varepsilon \quad \textbf{(1)}$$

DP :Price of Detached House Price per ㎡(Yen)

X_h :Main variables

X_1 : *GA* / Ground Area/Square Meters

X_2 : *FS* / Floor Space/Square Meters

X_3 : *RW* / Front Road Widths(10cm)

X_4 :*Age* /Age of Building

X_5 : *TS* / Time to Nearest Station

X_6 : *TT* / Travel Time to Central Buisiness District

Z_j :Other variables

Z_1 : *RT* / Market Reservation Time(week)

Z_2 : *OC* / other Characteristics

LD_k :Location(Ward) Dummy(k=0…K)

RD_l :Railway Dummy(l=0…L)

TD_m :Time Dummy(m=0…M)

V_n :Neighborhood Effect Variables

V_1 : *FR* / Floor Area Ratio/FAR

V_2 : *LR* / Lot Area Ratio/LAR

V_3 : *ZD* / Zoning Dummy

V_4 : *LU* / Land Use Condition

V_5 : *HC* / Household Characteristics

V_6 : *NOi* / Noise

u,v :Longitude,latitude

In the model and in accordance with previous studies, I set the following variables as major explanatory variables (X_h) that determine house prices: ground area (*GA*), floor space (*FS*), front road width (*RW*), age of building (*Age*), time to nearest station (*TS*), and travel time to the CBD (*TT*) (see, for example, Shimizu and Nishimura (2006), (2007)). The model composed of only these basic variables is referred to as Model-1.

In addition, I also considered some dummy variables including a location (ward) dummy (LD_j) and a railway dummy (RD_k) as locational characteristics as well as a time dummy (TD_l) to take into account the difference in the transaction time, in addition to the variables for the time elapsed between putting a house on the market to the conclusion of the contract (market reservation time, Z_1), and the direction the windows face (Z_2).

In addition to these variables, I expanded the model to one that takes neighborhood externalities into account, which is the major aim of this study. First, to handle neighborhood externalities explicitly, I incorporated variables (V_n) that represent neighborhood externalities.

V_n mainly comprises the following: zoning restrictions imposed by urban planning, which is generally available information, the local land-use conditions (*LU*), which can only be obtained by actual on-site observation, and the residential household characteristics (*HC*), as well as the road traffic noise (*NOi*), which is a strong environmental externality.

First, as variables related to land-use regulations, I set the following three variables that are easily obtained from publications and advertisements on the Internet: V_1: floor area ratio (*FR*), V_2: lot area ratio (*LR*), V_3: zoning dummy (*ZD*). The model in which these variables, which represent generally observable residential characteristics, are incorporated is referred to as Model-2.

Furthermore, I incorporated the following variables that can only be obtained by direct observation of properties: V_4: local land-use conditions (*LU*), V_5: household characteristics (*HC*), and V_6: road traffic noise (*NOi*). In addition, regarding small neighborhoods such as city blocks, while I assume that a group of households having the same preference forms a region as a result of spatial locational competition, I also assume that households having different preferences are located in different regions. In this case, the heterogeneity of variance between different regions is anticipated. Accordingly, when I assume that households having different preferences are located in different regions, to deal with the heterogeneity of variance, I incorporated fixed effects of the major variables (X_h). In this model, I incorporated the floor area ratio as a proxy variable representing land-use conditions and land-use density, as well as its cross terms with the major variables (X_h).

Even when control of the model by the incorporation of such neighborhood effect variables and fixing the effects of major variables are applied, it is still predicted that the problem of unobserved variables remains. To deal with this problem, I incorporated positional coordinate data (longitude and latitude) in the manner of Jackson (1979).

The model in which the neighborhood effect variables (V_n), the floor area ratio and its cross terms with the major variables to deal with the heterogeneity of variance, and the coordinate data (u, v) are incorporated is referred to as Model-3.

2.3.Difficult-to-observe neighborhood effect factors

In defining variables based on local land-use conditions (LU), I used individual building data from the "Survey on Land Use in Tokyo" in fiscal 2001. In this survey, data on 1,662,088 buildings are arranged as geographic information system (GIS) data (polygon data), wherein the conditions of use, area, structure, and other parameters of buildings are presented. With these data, I calculated (1) the number of buildings, (2) the average area of all buildings, and (3) the standard deviation of building area in each square of side length 500 m, which corresponds to the tallying region of the national census. The average area serves as a variable with the same characteristic as the building density, and the standard deviation of the building area was assumed as a proxy variable representing the appearance of the town. It is assumed that, in a region with a small standard deviation, houses have similar sizes along well-ordered streets, whereas in a region with a large standard deviation, the appearance of the town is not well-controlled.

As household characteristic (*HC*) data, I set the following items as variables in accordance with the national census: (1) the number of households, (2) the ratio of households with elderly people (the number of households having an individual aged 65 years or older/total number of households), (3) the ratio of households with an office worker (specialist and engineer employees + management-level employees + clerical employees/total households). The ratio of households with an office worker is assumed as a proxy variable representing the differences in the income and academic background in a region. This can be done because these employees are generally known to have a more academic background and higher income level on average than households with other occupations.

I also incorporated data on the floor space per household. The average area calculated from the "Survey on Land Use in Tokyo" is the average area of buildings existing in a mesh regardless of ownership, whereas the average area observed in the national census means the floor space attributed to a household, which is different from the former area index. I consider that, while the former is a proxy variable representing ambient environment, the latter can be used as a proxy variable representing assets or income.

In addition, I incorporated road traffic noise as a variable representing the quality of the local environment. Road traffic noise is not a cause of direct noise pollution; because a high degree of air pollution also occurs in a region with a high level of road traffic, I judged that road traffic noise is a representative index of local environmental quality. For the method used to measure road traffic noise, refer to the Appendix.

2.4. Data

2.4.1. Single-family house price

The analysis data used in this study is summarized in the following. As the subject of the analysis, I used the data for transactions concluded in the 23 wards of Tokyo (621 km2) within one year, from January to December 2000.

I used the prices of single-family houses and lands in "Residential Information Weekly" published by RECRUIT, Co. as the main information source. This magazine provides information on the characteristics and asking price of properties and land on a weekly basis, and includes the historical price data for individual properties from the time they are first placed in the magazine until they are removed because of sale or other reasons; the prices are renewed on a weekly basis. Regarding the price information available in the magazine, three items are listed: i) the initial asking price (first offer price) when the house was placed on the market, ii) the price upon removal from the magazine (estimated purchase price: first bid price), and iii) the actual transaction price, which is collected as a sample for statistical purposes. The first asking price represents the seller's desired price rather than the market value. In contrast, the transaction prices are enouph to estimate hedonic model considering neighborhood effects.

From the information published in Residential Information Weekly, I decided to use the price at the time of removal from the magazine upon the conclusion of the contract as the explained variable of the model. The price at the time of removal from the magazine is the first price offered by a prospective buyer; such a bid is offered through the process in which several particulars of characteristics and price are disclosed to the market via the magazine, and the price is decreased until the buyer responds to that information in a manner opposite to that of an auction. Thus, this price indicates the upper range of market price but it is extremely near transaction price. (Comparing resuls of 962 samples we observed between ii) last listed price and iii) transaction price, above the 95% samples are equal).

2.4.2. Data regarding house characteristics

The price of a single-family house is determined by information on the land and the building. As the numerical data representing the attributes of the land and building, I adopted ground area (*GA*), floor space (*FS*), and front road width (*RW*). The age of the building is the period from the construction of the building to the conclusion of the transaction. To take into account whether the house's windows are south-facing or not, I defined a south-facing dummy (*SD*). In addition, when a house has an adjacent private road, this is handled by a private-road dummy (*PD*); when a transaction is only for land without a building, it is handled by a land dummy (*LD*).

Furthermore, the convenience of public transportation from each house location is represented by the travel time to the CBD (*TT*) and the time to nearest station (*TS*). The travel time to the CBD is measured in the following way. First, I defined the CBD. The metropolitan area of

Tokyo is composed of 23 wards with Tokyo as its center containing a dense railway network. I designated terminal stations as the center of major business districts. The terminal stations are chosen to include six on the Yamanote Line: Tokyo, Shinagawa, Shibuya, Shinjuku, Ikebukuro, and Ueno, as well as Otemachi as the central station of the Tokyo Metro (Teito Rapid Transit Authority). Then, I investigated average travel times during the day from each station to the seven terminal stations, and set the minimum value as the travel time to the CBD for each property.

Regarding the time to the nearest station, times for different means of transportation are available. There are three means of transportation: by foot, by bus, and by car. However, the data for analysis only include houses within walking distance or bus-transportation distance. Therefore, any difference in the distance between the former and latter is controlled by a dummy variable (bus dummy: *BD*). In addition, the walking time (in minutes) is recorded when the house/land is within walking distance, and walking distance from the house/land to the bus stop and the onboard time from the bus stop to the nearest station (in minutes) are recorded in the case of houses/land in a bus-transportation area. The time to the nearest station (*TS*) is defined as (walking time to nearest station) + (walking time to bus stop) + (onboard time from bus stop to nearest station). Then, for a bus-transportation area, the cross term of the constant dummy variable with the time to the nearest station (*TS*) is incorporated in the bus dummy.

These variables are factors attributed to the location or buildings that are single-family houses; regional price differences are expected to exist as well. Therefore, I set a ward dummy (*WD*) to reflect differences in the quality of public services and differences in the prestige of regions. Furthermore, since most of the residential ground developments in the Tokyo metropolitan area have been carried out along railway lines, the price structure of houses differs along each railway line; therefore, I defined a railway line dummy (*RD*). Moreover, I used a time dummy (*TD*) to control differences in temporal price changes.

The sale price of each property is also affected by the fluidity and depth of the market. The time taken until the contract is concluded is considered to be affected by the period and location and by the level of activity of transactions in the market. I explain such market factors using variables such as market reservation time (*RT*). The RT is the time between the date when a land/house is placed on the market by a seller and the date when the transaction is concluded. Properties with a long market reservation time are regarded as having a price higher than the equilibrium price or participating in a thin market. Conversely, properties with a short market reservation time are regarded as being in a market with high fluidity or having a price close to or lower than the equilibrium price. For our purposes, I defined the market reservation time as the time between the first listing of the property in the magazine and its removal from the magazine due to contract conclusion.

Table 1 shows a list of the explanatory variables. The observation data consist of 13,822 transactions concluded between January and December 2,000.

Table 1. List of analyzed data

Symbols	Variables	Contents	Unit
GA	Ground Area/square meters	Ground Area	m^2
FS	Floor space/ square meters	Floor space.	m^2
RW	Front Road Widths	Front Road Widths	10cm
AGE	Number of years since construction	Period between the date when the data is deleted from the magazine and the date of construction of the building.	year
TS	Time to nearest station	Time distance to the nearest station (Time by Walk or Bus).	minute
TT	Travel Time to central business district	Minimum of railway riding time in daytime to Terminal 7 stations in 2000.	minute
RT	Market reservation time	Period between the date when the data appear in the magazine for the first time and the date of being deleted.	week
BD	Bus dummy	Whether the time distance includes riding time of bus 1, not including bus time including bus time 0.	(0,1)
FD	First floor dummy	The property is on the ground floor 1, on the other floors 0.	(0,1)
SD	South-facing dummy	Fenestrae facing south 1, other directions 0.	(0,1)
BU	Building characteristics	Building density-number of buildings,Average building floor space, in 500×500mesh.	—
CS	Census Data	Number of household ,etc	—
ZD	Zonning dummy	Land use zonning i 1, other land use 0.	(0,1)
LDk (*k=0,…,K*)	Location (Ward) dummy	*k th administrative district* 1, other district 0.	(0,1)

RDl (*l=0,…,L*)	Railway line dummy	*l th railway line* 1, other railway line 0.	(0,1)
TDm (*m=0,…,M*)	Time dummy (monthly)	*m th month* 1, other month 0.	(0,1)

2.4.3.Statistical distribution of single-family house prices

Table 2 shows descriptive statistics of the major variables.

The average single-family house price is 72.11 million yen, the minimum value is 10.50 million yen, and the maximum value is 398.00 million yen, with a fairly large standard deviation of 43.84 million yen. The data include a wide range of single-family houses from small dwellings to the so-called over-100-million-yen large ones. The average unit price is approximately 0.68 million yen/m^2 with a small variation.

The minimum ground area is 10.11 m^2 and the maximum area is 797 m^2, showing a large variation. The minimum floor space is 0 m^2 for land-only transactions and the maximum value is 649 m^2, with an average of 77 m^2.

Regarding the age of the buildings, while the average age is 5.28 years due to the inclusion of a large number of newly built houses, the maximum value is 41.33 years, with a right-skewed distribution.

Regarding the time to the nearest station, there are properties with a minimum value of 0 minutes that are located in front of a station; the maximum value is 36 minutes, with an average value of 10 minutes. On average, while many properties are conveniently located, some are located in areas with inconvenient public transport.

The minimum travel time to the CBD is 1 minute, which indicates that there are properties located adjacent to main terminal stations. The maximum travel time is 33 minutes, and the average time is 11 minutes. The variation is small because the subject of the analysis is the 23 wards of Tokyo, wherein which a dense railway network has been developed.

Table 2. Summary statististics of single-family house data.

Variables	Average	Standard Deviation	Minimum	Maximum
DP:Price ofDetached house price (10,000 Yen)	7,211.86	4,384.60	1,050.00	39,800.00
GA: Ground Area (㎡)	114.99	75.07	10.11	797.35
DP/*GA*	76.83	69.37	0.00	649.08
FS: Floor space (㎡)	4.85	2.48	2.00	50.00
RW:Front Road Widths	5.28	8.75	0.00	41.33
Age:Age of Building(monthes)	9.64	4.36	0.00	36.00
TS: Time to the nearest station: (minutes)	11.45	6.08	1.00	33.00
TT: Travel Time to Central Business District (minutes)	2.39	2.53	0.00	35.00
RT: Reservation Time (week)	68.67	24.08	12.00	392.00

2000/01-2000/12 n=12,954

3. Estimated results

3.1.Formulation of hedonic models

I formulated Model-1 on the basis of eq. (1) using the market for single-family houses and land in the 23 wards of Tokyo. The results are shown in the following.

$$\log DP/GA = \underset{(324.255)}{4.655} - \underset{(-42.569)}{0.001} \cdot GA + \underset{(15.276)}{0.001} \cdot FS + \underset{(19.749)}{0.016} \cdot RW - \underset{(-47.013)}{0.011} \cdot Age - \underset{(-20.938)}{0.010} \cdot TS - \underset{(-26.633)}{0.012} \cdot TT$$

$$- \underset{(-4.025)}{0.169} \cdot BD + \underset{(0.892)}{0.002} \cdot (BD \times \log TS) + \underset{(6.316)}{0.029} \cdot SD - \underset{(-35.761)}{0.284} \cdot LD - \underset{(-1.029)}{0.0007} \cdot MR + \beta_{1j} \sum_j WD_j$$

$$+ \beta_{2k} \sum_k RD_k + \beta_{3l} \sum_l TD_l + \varepsilon \qquad \textbf{(2)}$$

Adjusted R-square value: 0.645 (The numbers in parentheses indicate *t* values)
Number of observations: 12,954

Since the adjusted R-square value is 0.645, the formulated model has a fairly high explanatory power as a primitive model. The model was basically linear, and the selection of variables constituting the zoning dummy and railway line dummy was performed by a best-subset selection procedure based on Malow's CP when the major variables (X_h) were incorporated.

Using Model-1 as a starting model, I formulated other models wherein consideration is given to neighborhood effect variables (Table 3).

First, in Model-2, the urban-plan zoning dummies for house use, commercial use, and industrial use, as well as the floor area ratio and the lot area ratio were added. After incorporating the variables adopted in Model-1, I selected variables to be added by the best-subset selection procedure. As a result, the commercial-use zoning dummy and the floor area ratio were adopted.

Next, I formulated Model-3 by incorporating the ambient land-use conditions (*LU*), the household characteristics (*HC*) obtained in the national census, and the road traffic noise (*NOi*) as neighborhood effect variables, and by adding the cross terms of the floor area ratio with major variables as well as the latitude and longitude coordinates. After incorporating the variables adopted in Model-2, I included the squared terms and cubed terms of the latitude and longitude coordinates and added variables corresponding to V_4, V_5, and V_6. In concrete terms, I incorporated the variables adopted in Model-2 and the cross terms of the floor area ratio with the major variables and selected other variables by the best-subset selection procedure.

Table 3 shows the variables adopted. When the formulated models were compared, both Model-2 (adjusted R-Square: 0.693) and Model-3 (adjusted R-Square: 0.712) showed improved explanatory power as evaluated by the adjusted R-square value. In Model-2, the estimated values of the commercial-use zoning dummy and the floor area ratio have positive signs. However, neither estimate is statistically significant, even at the 5% significance level. When a comparison is made with Model-3, only when other variables such as neighborhood effect variables, position coordinates, and fixed effects are incorporated are the estimates of the commercial-use zoning dummy and the floor area ratio statistically significant at the 5% significance level and 1% significance level, respectively.

Regarding their signs, the commercial-use zoning dummy has the effect of decreasing the house price; we can see that house prices tend to be high in regions with a high floor area ratio. Since many regions have a poor living environment among commercial districts, I assume that such an effect was exhibited in the model.

Table3. Estimated results of modified hedonic models.

Dependent Variable

DP: Price of Detached House (in log) per ㎡

Independent Variables

	Model-2		Model-3	
***X*:Property Characteristics**	**Coefficient**	**t-value**	**Coefficient**	**t-value**
Constant	4.903	330.892	-48.523	-3.089
GA: Ground Area	-0.005	-50.466	-0.005	-42.660
FS: Floor space	0.001	24.264	0.001	12.686
RW:Front Road Widths	0.017	22.756	0.010	6.437
Age:Age of Building	-0.011	-47.183	-0.009	-19.166
TS: Time to the nearest station	-0.009	-20.190	-0.007	-7.637
BD: Bus Dummy	-0.141	-3.621	-0.134	-3.516
$TS \times BD$	0.002	0.623	0.002	0.692
TT: Travel time to the CBD	-0.011	-26.178	-0.007	-8.835
***Z*:Other Property Characteristics**	**Coefficient**	**t-value**	**Coefficient**	**t-value**
SD:South Face Dummy	0.027	6.511	0.029	7.190
LD:Land Dummy	-0.187	-24.146	-0.198	-26.244
MR:Market Reservation Time	-4.516E-04	-0.653	-4.383E-04	-0.653
V_1:Neiborhood Effects	**Coefficient**	**t-value**	**Coefficient**	**t-value**
ZD_1:*Zoning Dummy(Commercial use)*	0.005	0.555	-0.009	-1.055
ZD_2:*LR*(Lot area ratio)	5.E-06	0.153	2.017E-04	2.757
V_2:Neiborhood Effects by GIS	**Coefficient**	**t-value**	**Coefficient**	**t-value**
BU_1:*Average building area*	—	—	1.782E-03	12.752
BU_2:*Standard deviation of building area*	—	—	-1.576E-04	-4.912
BU_3:*BLD: Building Density*	—	—	1.560E-04	4.286
BU4:BLD^2	—	—	-6.659E-08	-3.858
CS_1:*Rate of Office worker*	—	—	6.414E-03	7.155
CS_2:*Average building area per household*	—	—	2.669E-03	12.314

EX:Noise	—	—		-1.858E-02	-4.139
Unobserved Spatial Effects(u,v)	**Coefficient**	**t-value**		**Coefficient**	**t-value**
latitude	—	—		3.384E-01	3.288
longitude[3]	—	—		4.658E-05	1.645
Fixed Effects by LR	**Coefficient**	**t-value**		**Coefficient**	**t-value**
GA: Ground Area	—	—		-3.940E-07	-1.425
FS: Floor space	—	—		9.231E-07	3.175
RW:Front Road Widths	—	—		1.696E-05	3.967
Age:Age of Building	—	—		-1.111E-05	-4.929
TS: Time to the nearest station	—	—		-1.084E-05	-2.379
TT: Travel time to the CBD	—	—		-1.061E-05	-3.071
Location (Ward) Dummy LD_j $(j=0,\ldots,J)$	**Yes**			**Yes**	
Railway/Subway Line Dummy RD_k $(k=0,\ldots,K)$	**Yes**			**Yes**	
Time Dummy TD_l $(l=0,\ldots,L)$	**Yes**			**Yes**	
Number of observations=	12,954			12,954	
Adjusted R-Square=	0.693			0.712	

3.2.Effect of neighborhood externalities on house price

Ambient land-use conditions*: *LU

According to Model-3 regarding ambient land-use conditions (*LU*), the estimated average building area per mesh (500 m x 500 m) had a positive sign and was statistically significant at the 1% significance level, indicating that an increase in the average building area positively affects the single-family house price. In contrast, the estimated standard deviation of the building area had a negative sign and was statistically significant. In addition, while the estimated building density had a positive sign, the estimated second-order term had a negative sign. In a region with a large average building area and uniform areas of buildings, the local environment is pleasant and the appearance of the town is well ordered, which causes house prices to increase. However, as the variation in building area increases, the local environment deteriorates, resulting in a decrease in house price.

Regarding building density, i.e., the degree of concentration of buildings in a regional unit, the primary term, which affects property values linearly, has a positive effect, whereas the estimated second-order term had a negative effect and was statistically significant. This suggests that an appropriate degree of building concentration has a positive effect on property values, whereas high concentration has a negative effect.

Household characteristics*: *HC

Regarding the household characteristics surveyed in the national census, the office worker ratio, as defined in section 2.3, and the average building area per household were adopted in Model-3. Both of these variables were estimated to have positive signs and were statistically significant at the 1% significance level. Since the academic background and income of people working in these occupations are expected to be high, the office worker ratio variable is considered to serve as a proxy variable representing both the ability to buy a house and the academic background. The average building area per household was adopted independently of the average building area (GIS data), and similarly it was estimated to be a positive externality.

In houses in a region with a large average building area according to GIS data, the likelihood that attention is paid to the façade (appearance) of the houses, such as tidy hedges, is high, so that an externality due to such physical contributions to the urban space is expected; I consider this the reason why the estimated average building area has a positive sign and is statistically significant. Meanwhile, at the level of the household unit, a large building area means a large area owned by a household; accordingly, the variable of the average building area shows characteristics of a proxy variable representing the assets and economic power of households. Thus, these two effects are estimated independently.

In other words, this suggests that in a region of households with high income and considerable assets, house prices tend to be high. Such a result is consistent with an economic model proposed by Rosen (1974), thus indicating that house prices are not determined only by physical factors such as the houses themselves.

Environmental externalities/road traffic noise*: *NOi

It is natural to expect that house prices are strongly affected by environmental externalities such as air cleanliness and noise, independent from the physical environment such as the town appearance set as *LU*. Here, I also incorporated the road traffic noise (*NOi*) as an environmental externality independent from ambient land-use conditions (*LU*) and household characteristics (*HC*). Since the road traffic noise is determined by the amount of road traffic, I considered that it could serve as not only the externality of noise but also as a proxy index representing air pollution and safety. Results of the estimation showed that road traffic noise had a negative sign and was statistically significant at the 1% significance level, and that an increase in noise by 1 dB decreased the house price by 1.8%.

Other factors

The fixed effect was controlled by the cross terms of the floor area ratio with the major variables, and the estimated values were statistically significant for all variables at the 1% significance level. Regarding the coordinate values (u, v), the cubes of latitude and longitude were adopted. This feature suggests that, in this model, geographical attributes remain that cannot be absorbed by the incorporation of travel time to the CBD, the land (ward) dummy, the railway line dummy, and the neighborhood effect factors.

3.3.Omitted variable bias

Through the above analyses, we have shown that neighborhood effect variables significantly affect house prices. Consequently, if we do not take these variables into consideration, omitted variable bias is generated in the values estimated by hedonic models. Therefore, I compared the three models formulated in this study, i.e., Model-1, Model-2, and Model-3 (Table 4), and measured the bias levels.

Comparison between Model-1 and Model-2

When Model-1 and Model-2 are compared, we see that the ground area (GA) is estimated as being too high in Model-1, but no difference is apparent in the estimated parameters of the other major variables between the two models. In concrete terms, while *ground area* was estimated as (-0.001) in Model-1, it was estimated as (-0.005) in Model-2.

The difference between these two models is due the presence or absence of the land-use regulation dummy and the floor area ratio. Since the floor area ratio and land-use regulation cause a change in the effectively available area, the occurrence of large differences between the statistics estimated using the models with and without these variables is expected.

For example, the ground area has a strong correlation with the effective area of land available for construction. Meanwhile, an increase in the floor area ratio increases the effective area available for construction in the vertical direction independent of the planar area; thus, there is a replacement relationship between the ground area and the floor area ratio. Consequently, when the floor area ratio and the commercial land-use regulation dummy were considered in Model-2, the estimated parameter showing a decrease in the house price accompanying an increase in the area was significant.

Regarding the variables other than *ground area,* it is difficult to assume a strong correlation with the floor area ratio or land-use regulations. Consistent with this hypothesis, the results showed no differences in the estimated parameters for the variables other than *ground area* between Model-1 and Model-2; thus, the correctness of this hypothesis has been verified.

Comparison between Model-2 and Model-3

Next, I compared Model-2 and Model-3. In Model-3 the variables *LU*, *HC*, and *NOi* are added and the fixed effects of the floor area ratio as well as the major variables and the coordinate data are incorporated. No changes were noted in the estimated values of ground area and floor space, but bias was generated for not only the variables representing locational characteristics such as the front road width, time to the nearest station, bus dummy, and the travel time to the CBD, but also for the age of the building.

The absolute values of the estimated parameters of Model-3 for these location characteristics and the age of the building were smaller than those of Model-2. In concrete terms, the estimated values were as follows:

Front Road Widths(*RW*)	: Model-2(0.017) / Model-3(0.010);	
Time to the nearest station (*TS*)	: Model-2(-0.009) / Model-3(-0.007);	
Bus dummy (*BD*)	: Model-2(-0.141) / Model-3(-0.134);	
Time to the nearest station (*TT*)	: Model-2(-0.011) / Model-3(-0.007);	and
Age of Building (*Age*)	: Model-2(-0.011) / Model-3(-0.009);	

This means that the effect of each variable considered as a neighborhood effect variable in Model-3 was integrated in the major variables in Model-2.

First, regarding *RW*, in residential districts in the 23 wards of Tokyo where houses are densely built, it is predicted that, as *RW* increases, insolation and ventilation are improved and public spaces are widened, leading to an improvement in the local environment. Therefore, when neighborhood externalities are not considered, it is possible that such environmental factors were absorbed in RW in the estimation.

Table4. Comparison of estimated parameters.

Method of Estimation

OLS

Dependent Variable

DP: Price of Detached House (in log) per ㎡

Independent Variables	**Model.1**		**Model.2**		**Model.3**	
***X*:Property Characteristics**	**Coefficient**	**t-value**	**Coefficient**	**t-value**	**Coefficient**	**t-value**
Constant	4.655	324.255	4.903	330.892	-48.523	-3.089
GA: Ground Area	-0.001	-42.569	-0.005	-50.466	-0.005	-42.660
FS: Floor space	0.001	15.276	0.001	24.264	0.001	12.686
RW:Front Road Widths	0.016	19.749	0.017	22.756	0.010	6.437
Age:Age of Building	-0.011	-47.013	-0.011	-47.183	-0.009	-19.166
TS: Time to the nearest station	-0.010	-20.938	-0.009	-20.190	-0.007	-7.637
BD: Bus Dummy	-0.169	-4.025	-0.141	-3.621	-0.134	-3.516
TS × *BD*	0.002	0.892	0.002	0.623	0.002	0.692
TT: Travel Time to CBD	-0.012	-26.663	-0.011	-26.178	-0.007	-8.835
V1:Neiborhood Effects	Yes		Yes		Yes	
V2:Neiborhood Effects by GIS	No		No		Yes	
Z:Other Property Characteristics	Yes		Yes		Yes	
Fixed Effects by LR	No		No		Yes	
Unobserved Spatial Effects	No		No		Yes	
Location (Ward) Dummy ***LDj (j=0,…,J)***	Yes		Yes		Yes	
Railway/Subway Line Dummy ***LDk (k=0,…,K)***	Yes		Yes		Yes	
Time Dummy ***TDl (l=0,…,L)***	Yes		Yes		Yes	
Number of observations=	12,954		12,954		12,954	
Adjusted R-Square=	0.645		0.693		0.712	

Regarding the variables representing the convenience of public transport such as *"time to nearest station"*, *"bus dummy"*, and *"travel time to the CBD"*, when environmental factors such as the neighborhood externalities were not considered, the house price was estimated to decrease with the physical distance from the CBD. However, as the distance from stations with large amounts of commerce (a large CBD) increases, environmental factors such as building density and road traffic noise improve. When this is taken into account, as the distance from the center of a city or from regions with a large amount of commerce increases, these environmental factors may cause the house price to increase. Thus, when the neighborhood externalities are taken into consideration, the effects of *"time to nearest station"*, *"bus dummy"*, and *"travel time to the CBD"* are less negative.

Regarding the age of buildings, it is unlikely that house prices decrease uniformly with age in the entire city. In districts where the environment is pleasant, the decrease in the house price with the time since the construction of the house is likely to be suppressed. Accordingly, by taking into consideration the neighborhood effect variables representing the local environment, the negative effect of age is reduced.

As described, we can see that, without taking into consideration the neighborhood effect variables, non-negligible omitted variable bias is generated in the hedonic model of the housing market in the 23 wards of Tokyo.

4. Concluding remarks

In this study, to estimate hedonic price functions, I focused on the problem of omitted variable bias and attempted to clarify the effect of the incorporation of neighborhood effect variables as a means of reducing this bias. In concrete terms, starting from a hedonic model wherein only the building attributes and locational characteristics are considered, I expanded the model by incorporating neighborhood effect variables and observed changes in the estimated parameters.

As a result of the analyses, the following items were clarified. Starting from Model-1, in which the estimation is carried out using only generally available information, I incorporated neighborhood effect variables into the model. The effects of incorporating the variables representing the land-use conditions (*LU*) were as follows.

-The estimated average building area in a mesh (500 m x 500 m) acted positively and was statistically significant. This may indicate that a large area per building acts as an externality, thereby affecting house prices positively. However, the estimated standard deviation had a negative sign and was statistically significant.

-As a reason for this, I found that, when the areas of buildings surrounding a house are similar, the region has a well-ordered appearance, which is a positive externality, whereas when the variation is large, the region does not appear well-ordered, leading to a negative externality.

-Regarding building density, while the estimated primary linear term had a positive sign, the estimated second-order term had a negative sign. An appropriate degree of building density produces a positive effect, but when the density becomes too high, a negative externality is induced.

Next, I considered the household characteristics among variables obtained in the national census. I also introduced road traffic noise as another environmental externality.

-The indices of the number of office workers and the average gross floor space per household were adopted, and their estimated values acted positively and were statistically significant. The number of office workers is a proxy variable representing the levels of academic background and income. The average building area per household is a proxy variable representing asset size. From this result, the importance of taking into consideration the individual attributes of households, such as income and academic background in the estimation using hedonic models, as shown by Rosen (1974), was also demonstrated.

-With respect to the road traffic noise, it was shown that an increase in the noise level of 1 dB decreases the property value by 1.8%.

Using these estimated results, I confirmed that omitted variable bias is generated in a model wherein neighborhood effect variables are not considered.

Starting from Model-1, the commercial-use zoning dummy and the floor area ratio were added in Model-2; the comparison between Model-1 and Model-2 showed that the estimated statistical value of the ground area in Model-1 was approximately twice that in Model-2. No effects of incorporating the variables of *"time to nearest station"*, *"bus dummy"*, *"travel time to the CBD"*, and *"age of building"* were noted.

In addition, when Model-3, in which the neighborhood effect variables were incorporated, was compared with Model-2, effects of *"time to nearest station"*, *"bus dummy"*, *"travel time to the CBD"*, and *"age of building"* were observed in Model-3.

-The estimated parameters in Model-2 compared with those in Model-3 were 1.7 times greater for the front road width and 1.3 times greater for the time to the nearest station, showing large differences. The absolute values of the estimated parameters of the bus dummy, the travel time to the CBD (*TT*), and the age of the building in Model-2 were larger than those in Model-3. This difference arises because, when the effects of the variables relating to the neighborhood effect variables are not considered as in Model-2, such effects are integrated in variables such as *"time to nearest station"*, *"bus dummy"*, *"travel time to the CBD"*, and *"age of building"*, so that these variables are estimated to have high values. Thus, the problem that the generation of omitted variable bias is at a non-negligible level was confirmed.

In this study, by considering the neighborhood effect variables represented by the environmental variables and the incomes of households in a small neighborhood unit, I aimed to expand the initial hedonic model and to clarify their effect on reducing omitted variable bias. As demonstrated by the series of estimated results, when the neighborhood effect variables were not considered in the estimation using hedonic models, bias was generated in the estimated parameters.

The application of hedonic models as an evaluation method for the effects of urban policies has been actively attempted in Japan. I hope that this study contributes to the evaluation of such policies.

Appendix. Measurement of road traffic noise

I measured the road traffic noise by field investigation and using a prediction method in accordance with ASJ Model 1998 by measuring and estimating equivalent noise levels LAeq,16h during the daytime reference time period (from 6 am to 10 pm). This method of measurement was chosen because I judged that the prediction would only yield the effects of roads on property values as the subject of the analysis, in contrast with actual measurements wherein measurement errors are generated due to variable factors (measurement day and time, measurement point). In addition, prediction errors were expected to be smaller than measurement errors.

The measurement sites comprised 202 sites in Setagaya ward, Tokyo. Measurements were performed at a point 1.2 m above the ground and 1-2 m from a building at the measurement site. At sites where a reflection effect was predicted, a correction of -2 dB was applied to the measured value. Measurements were performed between 2 pm and 6 pm on September 16 and 18, 2000, with a measurement period of 5 min at each site. The measurement procedure was as follows: signals (noise) received by an omnidirectional microphone (Rion NL-04 and NL-06) were recorded on a data recorder (Sony-D10). Then, an equivalent noise level LAeq.,5min for the 5 min signals, in which sounds other than the traffic noise on the subject road were excluded from the signals obtained from 10 min measurement, was obtained by an integral-average noise meter. This value was defined as the daytime equivalent noise level LAeq.,16h. We pressed the pause button to stop the analysis; then after removing the unwanted sounds, we restarted the analysis. Note that in the field investigation, in cases when measurement sites were at the back of a building facing another road, not only the traffic noise from the subject road but also traffic noise from vehicles passing the other road is included in the subject sound.

We used the road traffic noise, the amount of traffic of the road, and the distance from the road to produce a prediction model estimating the road traffic noise. The result of the estimation is as follows.

$$Noise_{it} = 46.167 + 0.832 \cdot (Rstr_{it} / Rdst_{it})$$

$Noise_{it}$: noise level at site i at time t
$Rstr_{it}$: amount of road traffic at site i at time t
$Rdst_{i}$: direct distance from site i to the nearest main road

Adjusted R-Square: 0.923(The number in the parentheses is a *t* value.)
Number of Observation: 202

This model has a strong explanatory power with an adjusted R-square value of 0.923, demonstrating that the model can be used as a prediction model.

Referenses

[1].Asami, Yasushi., Xiaolu Gao (2006) "Effect of Environmental Factors on Housing Prices: Application of GIS to Urban-Policy Analysis" Atsuyuki Okabe (ed.) GIS-Based Studies in the Humanities and Social Sciences, Taylor & Francis, Boca Raton, pp.211-228.

[2].Ekeland, I., J. J. Heckman and L. Nesheim, (2004), "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, Vol.112, pp.60-109.

[3].Jackson, J., (1979), "Intraurban Variation in the Price of Housing", *Journal of Urban Economics*, Vol 6(4), pp.464-479.

[4].Kanemoto, Y. and Nakamura, R (1986), "A New Approach to the Estimation of Structural Equations in Hedonic Models," *Journal of Urban Economics*, Vol.19, pp.218-233.

[5].Rosen, S., (1974), "Hedonic Prices and Implicit Markets, Product Differentiation in Pure Competition", *Journal of Political Economy*, Vol.82, pp.34-55.

[6].Shimizu, C. and K.G.Nishimura, (2006), "Biases in Appraisal Land Price Information: The Case of Japan," *Journal of Property Investment and Finance*, Vol.26, No.2, pp.150-175.

[7].Shimizu, C. and K.G.Nishimura, (2007), "Pricing structure in Tokyo metropolitan land markets and its structural changes: pre-bubble, bubble, and post-bubble periods," *Journal of Real EstateFinance and Economics*, Vol.35 (4), pp.475-496.

[8].Shimizu, C., K.G.Nishimura and K.Karato, (2007), "Nonlinearity of Housing Price Structure-Secondhand Condominium Market in Tokyo Metropolitan Area-," *CSIS Discussion Paper*: (The University of Tokyo), No.86.

[9].Shimizu, C., K.G.Nishimura and Y.Asami, (2004), "Search and Vacancy Costs in the Tokyo Housing Market: An Attempt to Measure Social Costs of Imperfect Information", *Review of Urban &Regional Development Studies*, Vol.16, No.3, pp.210-230.

[10].Tiebout, C.M., (1956), "A Pure Theory of Local Expenditures", *Journal of Political Economy*, Vol. 64(5), pp.416-424.

[11].Zhuang, Duan.,(2006) "Spatial Dependence and Neighborhood Effects in Mortgage Lending: A Geographically Weighted Regression Approach", *University of Southern California, Lusk Center for Real Estate-Working Papers*, 2006-1007.