Article

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

Chihiro Shimizu

Faculty of Economics and Business Administration, Reitaku University, 2-1-1 Hikarigaoka, Kashiwa, Chiba 277-8686, Japan; E-Mail:cshimizu@reitaku-u.ac.jp

Abstract: In the formulation of hedonic models, in addition to locational factors and building structures which affect the house prices, the generation of the omitted variable bias is thought to occur in cases when local environmental variables and the individual characteristics of house buyers 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, these variables have not been sufficiently considered in previous studies. We demonstrated that non-negligible levels of omitted variable bias are generated if these variables are not considered.

Keywords: hedonic model; identification; omitted variable bias; neighborhood effect; geographic information system

1. Introduction

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 impaired views. To cope with these problems, cities have actively adopted policies for preserving high-quality urban space through regulating land use and other measures. However, policies such as those on land-use regulations cannot always be achieved with the agreement of citizens due to uncertainties that arise about the effects of such policies. Accordingly, various attempts have been made to measure the effect of land-use regulation and other policies in terms of economic value. The most representative technique for such attempts is a hedonic model that focuses on property values.

However, several problems still remain in the formation of hedonic models [1]. The most significant problem is the identification of estimated variables. Namely, in estimations using a hedonic model, the 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) [2]; whereas in an actual city, heterogeneity of family income should be taken as a prerequisite [3]. 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 area, 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 have been explicitly incorporated [4]. The lack of such studies 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 [1].

When there are unobservable variables such as these, what is the actual extent of the bias that exists in the estimated coefficient? If the extent of the bias were such that it could be ignored, then it would not be necessary to pay that much attention to the problem of omitted variable bias when estimating hedonic functions. The aim of this paper is to clarify the extent of the omitted variable problem that accompanies the existence of unobservable variables.

In terms of methods for addressing the problem of omitted variable bias accompanying unobservable variables, attempts have been made to estimate hedonic functions by performing market segmentation [5,6] or applying a spatial econometric method incorporating spatial autocorrelation in the error term. However, if segmenting regions, while various parameters are estimated in a more region-based form, it is difficult to obtain consistency across adjacent regions due to the segmentation of markets.

Another method is to apply spatial econometrics, which has been developing rapidly in recent years. In terms of methods that consider variables for which spatial observation is not possible, parametric and semi-parametric methods have traditionally been proposed.

As a representative parametric method, an estimation method has been proposed that aims to increase flexibility of fit by means of a high-dimensional polynomial equation using coordinate values (latitude, longitude). This is the so-called Parametric Polynomial Expansion model proposed by Jackson [7], which inputs the squares and cubes of coordinate values and a multi-dimensional cross term into explanatory variables. However, a highly collinear relationship is to be expected between the squares and cubes of coordinate values and the multi-dimensional cross term, and increasing the polynomial equation dimensions in order to increase flexibility of fit leads to a reduction in the degree of freedom. As a result, reduced probability with respect to the degree of confidence of individual estimate values is a concern.

With regard to semi-parametric methods, a Geographically Weighted Regression (GWR) model has been proposed that smooths unobservable geographic differences with coordinate values only, with the purpose of eliminating the effect of those differences on estimated values. Semi-parametric methods are convenient in that they do not assume a function form in advance with respect to the hedonic model. In addition, while it is necessary to introduce multiple geographic attributes to capture geographic local changes with parametric methods, semi-parametric estimation involving local regression performs correction using only coordinate values. However, since complex numerical calculations are performed,

the estimation procedure requires an enormous amount of time. Estimation methods that take spatial correlation of the error term into account have also been proposed.

However, all of these approaches do no more than *indirectly* resolve the problem of unobservable variables by devising estimation methods.

Resolving this kind of problem *directly* means addressing it by, insofar as possible, creating observable variables for variables that are considered unobservable through devising GISs and the like. With this in mind, this paper will clarify the extent of the omitted variable problem.

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

2. Hedonic Model with Consideration Given to Neighborhood Effect

2.1. Estimation Model

The following hedonic model was set up using the market of single-family houses and land.

$$\log DP/GA = a_0 + \sum_h a_{1h} X_h + \sum_i a_{2i} Z_i + \sum_j a_{3j} \cdot LD_j + \sum_k a_{4k} \cdot RD_k$$
$$+ \sum_l a_{5l} \cdot TD_l + \sum_m a_{6m} V_m + a_7 u + a_8 v + \varepsilon \tag{1}$$

DP: Price of Detached House Price per Square Meter (Yen)

X_h: Main Variables

$\quad X_1$ (GA): Ground Area (Square Meters)

$\quad X_2$ (FS): Floor Space (Square Meters)

$\quad X_3$ (RW): Front Road Widths (10cm)

$\quad X_4$ (Age): Age of Building

$\quad X_5$ (TS): Time to Nearest Station

$\quad X_6$ (BD): Bus Dummy

$\quad X_7$: $X_5 \times X_6$

$\quad X_8$ (TT): Travel Time to Central Business District

Z_i: Other Variables

$\quad Z_1$ (RT): Market Reservation Time (week)

$\quad Z_2$ (SD): South-facing Dummy

$\quad Z_3$ (LaD): Land Dummy

LD_j: Location (Ward) Dummy (j = 0…J)

RD_k: Railway Dummy (k = 0…K)

TD_l: Time Dummy (l = 0…L)

V_m: Neighborhood Effect Variables

V_1 (*FR*): Floor Area Ratio/FAR

V_2 (*ZD*): Zoning Dummy

V_{3-6}: LU_n/Land Use Condition

    V_3 (LU_1): *Average Building Area*

    V_4 (LU_2): *Standard Deviation of Building Area*

    V_5 (LU_3): *Building Density*

    V_6 (LU_4): *Building Density^2*

V_{7-8}: HC_o/Household Characteristics

    V_7 (HC_1): *Rate of Office Worker*

    V_8 (HC_2): *Average Building Area per Household*

    V_9 (*EX*): Noise

u, v: Longitude, Latitude[3]

In the model and in accordance with previous studies, the following variables were set as major explanatory variables (*Xh*) that determine house prices: ground area (*GA*), floor space (*FS*), front road width (*RW*), age of building (*Age*), time to nearest station (*TS*), bus dummy (*BD*), cross term of *TS* and *BD* (*TS* × *BD*), and travel time to the CBD (*TT*) (see, for example, Shimizu and Nishimura, Diewert and Shimizu [8–10]). In addition, some dummy variables were considered including a location (ward) dummy (*LDj*) and a railway dummy (*RDk*) as locational characteristics as well as a time dummy (*TDl*) to take into account the difference in the transaction time, on top of the variables for the time elapsed between putting a house on the market to the conclusion of the contract (market reservation time, *Z1*), the direction the windows face (*Z2*), and a dummy variable indicating that a transaction is made only for land (land dummy, *Z3*). The model composed of only these basic variables is referred to as Model-1.

In addition to these variables, the model was expanded to one that takes the neighborhood effect into account, which is the major aim of this study. First, to handle the neighborhood effect explicitly, variables (*Vm*) were incorporated that represent the neighborhood effect [11].

Vm mainly comprises the following: zoning restrictions imposed by urban planning, which is publicly available information, the local land-use conditions (*LUn*), which can only be obtained by actual on-site observation, the residential household characteristics (*HCo*), and environmental externality (*EX*).

First, as variables related to land-use regulations, the following two variables were set that are easily obtained from publications and advertisements on the Internet: V1: floor area ratio (*FR*) and V2: zoning dummy (*ZD*). The model in which these variables, which represent generally observable residential characteristics, are incorporated is referred to as Model-2.

Furthermore, the following variables, which can only be obtained by direct observation of properties: V3 to V6: local land-use conditions (*LUn*), and V7 to V8: household characteristics (*HCo*), were incorporated.

Even when the model is controlled by the incorporation of such neighborhood effect variables, it is still predicted that the unobserved variables problem remains. To deal with this, geographical coordinate data (longitude and latitude) based on Jackson's method [5] was incorporated.

The model in which the neighborhood effect variables (*Vm*) and the coordinate data (*u, v*) are incorporated is referred to as Model-3.

2.2. Data

2.2.1. Single-Family House Price

The analysis data used in this study is summarized in the following data for transactions concluded in the 23 wards of Tokyo (621 km^2) within one year, from January to December 2000.

The prices of single-family houses and land in "Residential Information Weekly" published by RECRUIT Co., as the main information source, were used. This weekly magazine provides information on the characteristics and asking price of properties and land and includes the historical price data for individual properties, from the time the properties are first listed in the magazine until they are removed because of sale or other reasons. The prices are refreshed on a weekly basis. Three points of price information are available in the magazine: (1) the initial asking price (first offer price) when the house was placed on the market, (2) the price upon removal from the magazine (estimated purchase price: first bid price), and (3) 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 enough to estimate the hedonic model considering neighborhood effects.

From the information published in Residential Information Weekly, it was decided to use the price at the time of removal from the magazine as a result of contract conclusion, 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 a process in which several particulars of the 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 that is opposite to that of an auction. Thus, this price indicates the upper range of the market price but it is extremely near the transaction price. (Comparing results of 962 samples we observed between (2) last listed price and (3) transaction price, more than 95% of the samples were equal).

2.2.2. Data Regarding House Characteristics

The price of a single-family house is determined by information on the land and the building. Ground area (*GA*), floor space (*FS*), and front road width (*RW*) were adopted as the numerical data representing the attributes of the land and building. 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, a south-facing dummy (*SD*) was defined. In addition, when a transaction is only for land without a building, it is handled by a land dummy (*LaD*).

Furthermore, the convenience of public transportation from each house location is represented by the travel time to the "Central Business District" or the CBD (*TT*) and the time to nearest station (*TS*). The travel time to the CBD is measured in the following way. First, the CBD was defined. The metropolitan area of Tokyo is composed of 23 wards with Tokyo as its center containing a dense railway network.

Terminal stations were designated as the center of major business districts. The terminal stations chosen included 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, the average travel times during the day from each station to the seven terminal stations were investigated, and the minimum value as the travel time to the CBD for each property were set.

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 riding time from the bus stop to the nearest station (in minutes) are recorded for 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) + (riding 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 single-family houses; regional price differences are expected to exist as well. Therefore, a location (ward) dummy (*LD*) was set to reflect differences in the quality of public services and differences in area prestige. 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, a railway line dummy (*RD*) was defined. Moreover, a time dummy (*TD*) was used 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 transaction activity in the market. These kinds of market factors are explained 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 the purpose of this study, the market reservation time was defined as the time between the first listing of the property in the magazine and its removal from the magazine as a result of contract conclusion.

2.2.3. Data on Neighborhood Effect Factor

In defining variables based on local land-use conditions (*LUn*), individual building data was used 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) presenting the conditions of use, area, structure, and other parameters of buildings. With these data, (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, were calculated. 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 town's appearance is not well-controlled.

As household characteristic (*HCo*) data, the following items were set as variables in accordance with the national census: 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.

Data on the floor space per household was also incorporated. 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. It is considered that, while the former is a proxy variable representing ambient environment, the latter can be used as a proxy variable representing assets or income.

Table 1 shows a list of the explanatory variables. The observation data consist of 13,822 transactions concluded between January and December 2000.

Table 1. List of explanatory variables.

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	10 cm
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.	Years
TS	Time to nearest station	Time distance to the nearest station (Time by Walk or Bus)	Minutes
BD	Bus dummy	Whether the time distance includes riding time of bus 1	(0,1)
		Not including bus time 0	
TT	Travel Time to central business district	Minimum railway riding time in daytime to seven terminal stations in 2000	Minutes
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.	Weeks
SD	*South-facing dummy*	Windows facing south 1	*(0,1)*
		other directions 0	
LaD	Land dummy	Transaction is only for land without building 1	(0,1)
		otherwise 0	
LD_j *(j = 0,...,J)*	Location (Ward) dummy	*j th administrative district* 1	(0,1)
		other district 0	
RD_k *(k = 0,...,K)*	Railway line dummy	*k th railway line* 1	(0,1)
		other railway line 0	
TD_l *(l = 0,...,L)*	Time dummy (quarterly)	*l th quarter* 1	(0,1)
		other quarter 0	
FR	Floor Area Ratio	Floor area ratio	%

Table 1. *Cont.*

Symbols	Variables	Contents	Unit
ZD	Zoning dummy	Land use zoning i 1	(0,1)
		other land use 0.	
LUn (n = 1,2,3,4)	Land use condition	(1) Average building area, (2) Standard deviation of building area, (3) Building density, and (4) Building density squared	
HCo (o = 1,2)	Household Characteristics	(1) Rate of office worker and (2) Average building area per household	
EX	Road traffic noise	Road traffic noise	db
u, v	Longitude, Latitude	Geographical coordinates	

2.2.4. Statistical Distribution of Single-Family House Prices

Table 2 shows descriptive statistics of the major variables.

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

Variables	Average	Standard Deviation	Minimum	Maximum
DP: Price of Detached house price (10,000 Yen)	7211.86	4384.60	1050.00	39,800.00
GA: Ground Area (m^2)	114.99	75.07	10.11	797.35
DP/*GA*	68.67	24.08	12.00	392.00
FS: Floor space (m^2)	76.83	69.37	0.00	649.08
RW: Front Road Widths	4.85	2.48	2.00	50.00
Age: Age of Building(years)	5.28	8.75	0.00	41.33
TS: Time to the nearest station (minutes)	9.64	4.36	0.00	36.00
TT: Travel Time to Central Business District (minutes)	11.45	6.08	1.00	33.00
RT: Reservation Time (week)	2.39	2.53	0.00	35.00
01/2000–12/2000	n = 12,954			

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 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 a dense railway network has been developed.

3. Estimated Results

3.1. Formulation of Hedonic Models

Model-1 was formulated on the basis of equation (1) using the market for single-family houses and land in the 23 wards of Tokyo. The results are shown in the first column of Table 3.

Table 3. Estimated results of modified hedonic models.

Independent Variables	Model-1		Model-2		Model-3	
*X:*Property Characteristics	Coefficient	*p*-value	Coefficient	*p*-value	Coefficient	*p*-value
Constant	4.7131	***	4.6550	***	−53.42	***
GA: Ground Area	−0.0014	***	−0.0014	***	−0.0015	***
FS: Floor space	0.0008	***	0.0008	***	0.0007	***
RW: Front Road Widths	0.0169	***	0.0157	***	0.0143	***
Age: Age of Building	−0.0115	***	−0.0115	***	−0.0116	***
TS: Time to the nearest station	−0.0105	***	−0.0099	***	−0.0100	***
BD: Bus Dummy	−0.1773	***	−0.1691	***	−0.1622	***
TS × *BD*	0.0028		0.0024		0.0029	
TT: Travel Time to CBD	−0.0128	***	−0.0118	***	−0.0100	***
Z:Other Property Characteristics	Coefficient		Coefficient		Coefficient	
RT: Market Reservation Time	−0.0006		−0.0008		−0.0007	
SD: South Face Dummy	0.0292	***	0.0285	***	0.0301	***
LaD: Land Dummy	−0.2824	***	−0.2836	***	−0.2932	***
Location (Ward) Dummy $LD_j (j = 0,…,J)$	Yes		Yes		Yes	
Railway/Subway Line Dummy $RD_k (k = 0,…,K)$	Yes		Yes		Yes	
Time Dummy $TD_l (l = 0,…,L)$	Yes		Yes		Yes	

Table 3. *Cont.*

Independent Variables	Model-1		Model-2		Model-3	
V1-2: Neighborhood Effects	Coefficient		Coefficient		Coefficient	
FR: Floor area ratio	—	—	0.0002	***	0.0003	***
ZD: Zoning Dummy (Commercial use)	—	—	−0.0056		−0.0148	*
V3-8:Neighborhood Effects by GIS	Coefficient		Coefficient		Coefficient	
LU1: Average building area	—	—	—	—	0.0018	***
LU2: Standard deviation of building area	—	—	—	—	−0.0001	***
LU3: BLD: Building Density	—	—	—	—	0.0001	***
LU4: BLD²	—	—	—	—	0.0000	***
HC1: Rate of Office worker	—	—	—	—	0.0058	***
HC2: Average building area per household	—	—	—	—	0.0029	***
EX: Road traffic noise	—	—	—	—	−0.0160	***
latitude	—	—	—	—	0.4028	***
longitude³	—	—	—	—	0.0000	
Number of observations =	12,954		12,954		12,954	
Adjusted R-Square =	0.644		0.645		0.662	
AIC =	−3327		−3372		−4000	

Notes: Method of Estimation: OLS(Ordinary Least Squares regression); Dependent Variable: *DP*: Price of Detached House (in log) per m²; Signs ***, **, and * represent the estimated parameters are significant at 1%, 5% and 10% respectively.

Since the adjusted R-square value is 0.644, 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 (*Xh*) were incorporated.

Using Model-1 as a starting model, other models were formulated wherein consideration was given to neighborhood effect variables (Model-2 and Model-3 in 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 were added. After incorporating the variables adopted in Model-1, variables were selected 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, Model-3 was formulated by incorporating the ambient land-use conditions (*LU*), and the household characteristics (*HC*) obtained in the national census, and the latitude and longitude coordinates. After incorporating the variables adopted in Model-2, the cubed term of the latitude and the linear term of longitude coordinates were included and variables corresponding to V3 to V8 were added.

Here, with the expansion of the estimation model, we verified the correlation coefficient between the key variables Ground Area (*GA*), Floor Space (*FS*), Front Road Width (*RW*), Age of Building (*Age*), Time to the Nearest Station (*TS*), and Travel Time to CBD (*TT*) and the expanded variables for Model 2 and Model 3; the results showed that between the six key variables and the added variables, there were none which had a significant correlation coefficient. As a result, the problem of multicollinearity does

not occur. In addition, the comparison here was made using a simple linear model, rather than assuming a non-linear model as in Shimizu, Nishimura, and Karato [12].

3.2. Effect of Neighborhood on House Price

3.2.1. Ambient Land-Use Conditions: LU

According to Model-3 regarding ambient land-use conditions (*LU*), the estimated average building area per mesh (500 m × 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. A simple computation shows that as the building density increases, the house price increases, until the 95% of the building density value in the sample, at which point the house price decreases. 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 value, whereas high concentration has a negative effect.

3.2.2. Household Characteristics: HC

Regarding the household characteristics surveyed in the national census, the office worker ratio, as defined in Section 2.2.2, 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; this was considered to be 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 [2], thus indicating that house prices are not determined only by physical factors such as the houses themselves.

3.2.3. Other Factors

The coordinate values (*u*, *v*), the cubed term of latitude and the linear term of 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, a comparison was made of the estimated parameters of main variables in the three models formulated in this study, *i.e.*, Model-1, Model-2, and Model-3 (Table 4), and the bias levels were measured. The presence of omitted variable bias is confirmed by comparing the estimated parameters in base models to the confidence intervals of estimated parameters in modified models.

Table 4. Comparison of the estimated parameters among three models.

Estimated Parameters	Model-1	Model-2	Test *	Model-2	Model-3	Test *
GA: Ground Area	−0.00143	−0.00139		−0.00139	−0.00147	b
FS: Floor space	0.00077	0.00077		0.00077	0.00070	
RW: Front Road Widths	0.01693	0.01575		0.01575	0.01432	c
Age: Age of Building	−0.01146	−0.01148		−0.01148	−0.01162	
TS: Time to the nearest station	−0.01052	−0.00994		−0.00994	−0.01002	
BD: Bus Dummy	−0.17730	−0.16906		−0.16906	−0.16220	
TT: Travel Time to CBD	−0.01279	−0.01177	b	−0.01177	−0.01003	a
Neighborhood Effects	No	Yes		Yes	Yes	
Neighborhood Effects by GIS	No	No		No	Yes	
Unobserved Spatial Effects	No	No		No	Yes	

* The test verified whether or not the estimate values of right column are within the 99%, 95%, and 90% confidence intervals for the parameters estimated for left column; "a", "b", "c" indicate the estimated parameters in the left column are not in the 99%, 95%, and 90% confidence intervals, respectively, of the estimated parameters in the right column.

3.3.1. Comparison between Model-1 and Model-2

When Model-1 and Model-2 are compared, we see that the "travel time to the CBD" is estimated lower in absolute value in Model-2, but no difference is apparent in the estimated parameters of the other major variables between the two models.

Specifically, in Table 4, we verified whether or not the Model-2 estimate values are within the 99%, 95%, and 90% confidence intervals for the parameters estimated for Model-1. The results show that only "Travel Time to CBD" is not within the 95% confidence interval.

The difference between these two models is caused by 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. However, a clear difference is confirmed only in the estimated parameter of "travel time to the CBD" and not in other variables such as "ground area" and "front road width", which are thought to correlate with neighborhood effect variables such as "floor area ratio" and "land use regulations".

This result does not necessarily indicate the absence of the problem of biases in the estimated parameters induced by omitting neighborhood effect variables.

Thus we continue to incorporate more specific neighborhood effect variables measured by GIS (Model-3), rather than just incorporating those variables easily obtained from publication such as "floor area ratio" and "land use regulations" (Model-2).

3.3.2. Comparison between Model-2 and Model-3

Next, Model-2 and Model-3 were compared. In Model-3 the variables LU, HC, and NOi are added and the coordinate data are incorporated. No clear changes were noted in the estimated values of building characteristics such as "floor space", "age of building", and locational characteristics such as "time to the nearest station" and "bus dummy" which are thought to correlate with housing characteristics. However a certain bias was observed for locational characteristic variables such as "ground area" and "front road width", as well as "travel time to the CBD".

In Table 4, in a similar manner to the comparison above, we verified the level of probability that the Model-3 estimation parameters would not fall within the confidence intervals for the Model-2 estimation parameters. This showed that the estimated parameters did not fall within the 90% confidence interval for "Ground Area", the 90% confidence interval for "Front Road Width", and the 99% confidence interval for "Travel Time to CBD". In other words, these show the probability that there will be a difference for the respective estimated values.

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 "front road width (*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 effects are not considered, it is possible that such environmental factors were absorbed in RW in the estimation.

Regarding the variables representing the convenience of public transport such as "travel time to the CBD", when environmental factors and household characteristics 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 commercial areas (a large CBD) increases, environmental factors such as building density 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 effect and household characteristics are taken into consideration, the effect of "travel time to the CBD" is less negative.

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. Conclusions

In this study, to estimate hedonic price functions, a focus was placed on the problem of omitted variable bias and an attempt was made to clarify the effect of incorporating 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, the model was expanded by incorporating neighborhood effect variables and observed changes in the estimated parameters.

By considering the neighborhood effect variables represented by the environmental variables and the incomes of households in a small neighborhood unit, the objective was to expand the initial hedonic model and to clarify the 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 Green buildings has been actively attempted in Japan. We hope that this study contributes to the evaluation of such policies.

Acknowledgments

Conflicts of Interest

The author declares no conflict of interest.

References

1. Ekeland, I.; Heckman, J.J.; Nesheim, L. Identification and Estimation of Hedonic Models. *J. Polit. Econ.* **2004**, *112*, 60–109.
2. Rosen, S. Hedonic Prices and Implicit Markets, Product Differentiation in Pure Competition. *J. Polit. Econ.* **1974**, *82*, 34–55.
3. Kanemoto, Y.; Nakamura, R.A New Approach to the Estimation of Structural Equations in Hedonic Models. *J. Urban Econ.***1986**, *19*, 218–233.
4. Asami, Y.; Gao, X.L. Effect of Environmental Factors on Housing Prices: Application of GIS to Urban-Policy Analysis. In *GIS-Based Studies in the Humanities and Social Sciences*, Okabe, A., Ed.; Taylor & Francis: Boca Raton, FL, USA, 2006; pp.211–228.
5. Goodman, A.C.; Thibodeau, T.G. Dwelling Age Heteroskedasticity in Repeat Sales House Price Equations. *Real Estate Econ.* **1998**, *26*, 151–171.

6.   Goodman, A.C.; Thibodeau, T.G. Housing market segmentation and hedonic prediction accuracy. *J. Hous. Econ.* **2003**, *12*,181–201.

7.   Jackson, J. Intraurban Variation in the Price of Housing. *J. Urban Econ.* **1979**, *6*, 464–479.

8.   Shimizu, C.; Nishimura, K.G. Biases in Appraisal Land Price Information: The Case of Japan. *J. Prop. Invest. Finance* **2006**, *26*, 150–175.

9.   Shimizu, C.; Nishimura, K.G. Pricing structure in Tokyo metropolitan land markets and its structural changes: Pre-bubble, bubble, and post-bubble periods. *J. Real Estate Finance Econ.* **2007**, *35*, 475–496.

10. Diewert, E.; Shimizu, C. Residential Property Price Indexes for Tokyo. Available online: http://www.economics.ubc.ca/files/2014/04/pdf_paper_erwin-diewert-13-7-residential-property. pdf (accessed on 28 April 2014)

11. Shimizu, C.; Nishimura, K.G.; Asami, Y. Search and Vacancy Costs in the Tokyo Housing Market: An Attempt to Measure Social Costs of Imperfect Information. *Rev. Urban Reg. Dev. Stud.* **2004**, *16*, 210–230.

12. Shimizu, C.; Nishimura, K.G.; Karato, K. Nonlinearity of Housing Price Structure—The Secondhand Condominium Market in Tokyo Metropolitan Area. Available online: http://www.csis.u-tokyo.ac.jp/dp/86r1.pdf (accessed on 1 May 2010).