Department of Social Systems and Management

Discussion Paper Series

No.1266

Development of Algorithms for Estimating Apartment Rents in Metropolitan Area Based on a Combined Micro-Macro Approach

by

Jia-Ping HUANG, Ushio SUMITA,
Ryo TERAKADO, Takenori YANAGAWA, and Jun YOSHII

March 2011

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Development of Algorithms for Estimating Apartment Rents in Metropolitan Area Based on a Combined Micro-Macro Approach

Jia-Ping Huang Ushio Sumita Ryo Terakado

Takenori Yanagawa Jun Yoshii*

Graduate School of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, JAPAN

February 2011

Abstract

Information about the rental costs of the large apartment buildings is asymmetric in that the real estate companies tend to disclose such information to a customer only for the large apartment buildings of potential interest to the customer, and any information irrelevant to the ongoing business would be sealed. This information imbalance prevents the market to be transparent, and the economic market principles are often ignored. In order to overcome this pitfall, this paper aims at developing a numerical algorithm for estimating the unit rent of a large apartment building based on a set of real data in the metropolitan Tokyo. The algorithm is based on the combined micro-macro approach, where the local information such as the nearest rail station, the distance to it, and the like would be used first to estimate the unit rent through the micro approach. For the macro approach, the linear regression is employed based on the real data, and the resulting estimation formula would yield the second estimate. The two estimates would then be linearly combined, where the optimal weighting factor would be determined so as to minimize the discrepancy between the combined estimated values and the unit rents obtained from the real data.

Keywords Estimation, apartment rent, metropolitan area, micro-macro approach

1 Introduction

Concerning rental prices of large apartment buildings, a structural information asymmetry exists between real estate companies and customers in that a typical real estate company would disclose only partial information to its customers. In other words, it is often difficult for a customer to acquire the knowledge about rental prices of large apartment buildings in a broader market, since any real estate company would inform the customers of only the rental prices of the large apartment buildings for which the customers show an interest. It would be virtually impossible for the customer to confirm the appropriateness of the rental price in the entire regional market. The customer often has no choice but to accept the current rental price at the time of contract renewal. This information asymmetry results in the obstruction of the market transparency, destroying the conditions needed for fair trades in a perfect market.

*Corresponding author, E-mail: yoshii40@sk.tsukuba.ac.jp

In order to overcome this pitfall, the extensive literature exists concerning how to estimate hedonic prices of the real estate properties. To the best knowledge of the authors, this type of research can be traced back to 1970's represented by the original paper by Rosen (1974), where the theory of hedonic prices was formulated as a problem in the economics of spatial equilibrium in which the entire set of implicit prices would guide both consumer and producer locational decisions. Empirical implications were obtained via hedonic price regression. The analysis has been extended by Goodman (1978), highlighting differential effects of city v.s. suburb as well as structural and neighborhood. The reader is referred to a succinct summary by Sheppard (1999) and an excellent survey paper by Boyle & Kiel (2001). Concerning prices of real estate properties, some related papers include Bailey et al. (1963) establishing a price index via regression, Case & Quigley (1991) estimating housing prices by combining repeat sales of unchanged properties and improved properties together, Quigley (1995) combining hedonic and repeat sales methods in a unified manner, Hwang & Quigley (2009) by demonstrating the danger of exclusive reliance on analytical models through empirical study and simulation, and Rosenthal (2009) developing an elegant analytical model to cope with a deadline to sell the home, a common feature of the housing market, beyond which fixed and variable penalty costs might be imposed to both the homeowner and selling agent, among others. Empirical studies have been also developed, represented by Poon (1978) focusing on railway pollution in London, Canada, Debrezion et al. (2007) incorporating accessibility variables such as highways and railways, Do & Grudnitski (1995) analyzing the impact of golf course properties, Tyrväinen & Miettinen (2000) considering residents' valuations attached to forests, Morancho (2003) analyzing the link between housing prices and urban green areas endowments, Conroy & Milosch (2009) estimating the coastal premium, and Brunauer et al. (2009) assessing the spatially heterogeneous structure of price gradients in Vienna, among others.

The purpose of this paper is to propose a new comprehensive scheme for estimating rental prices of large apartment buildings based on a set of real data in the metropolitan Tokyo. As for key factors to determine the rental price of a large apartment building, we categorize them into two classes. One class of such factors would describe the locational condition around the apartment building, represented by the nearest rail station and the distance to it. The other class would consist of the features of the apartment building, including the area, the age, the structural type, the construction cost per m^2, and the like. In order to assess the impact of the former class on the unit rent per day $\cdot$ m^2, a micro approach may be effective where local information would be analyzed locally, say among apartment buildings within a vicinity of a rail station. For the latter class, a macro approach may be useful based on statistical regression. Since the unit rent is affected jointly by the key factors in the two classes, it would be better off to utilize both the micro approach and the macro approach in a combined manner. In Jin, Huang, Sumita and Lu (2008), a tentative approach is proposed for estimating the unit rent of a small apartment building, where a linear combination of the estimated value based on a micro approach and that on a macro approach is employed. In this paper, we follow this line of research and deal with unit rents of large apartment buildings. This task is more difficult in that the current unit rent of an apartment building must be derived from the balanced sheet of the real estate company for large apartment buildings, while this information is directly available for small apartment buildings.

In this paper, we first decompose the data set into subgroups along two axes so as to reduce the variance of the unit acquired prices of large apartment buildings paid by the real estate company. The first axis is the unit acquired price itself, decomposing the price range into 4 intervals. The second axis is the regional characteristics of the 23 wards in the metropolitan Tokyo, grouping them into 3 geographical regions: one with concentration of expensive large apartment buildings, another having large apartment buildings of low unit acquired prices as a majority, and the third between the two. The combined micro-macro approach would be implemented separately in each of $4 \times 3 = 12$ subgroups.

A micro approach is built upon a new concept of "the value of a rail station," characterized by the mean and the variance of the unit rents of the large apartment buildings having the rail station as the nearest in common. For a large apartment building having the same rail station as the nearest, the unit rent per day $\cdot$ m^2 can then be estimated based on the value of the rail station and the distance between the apartment building and the rail station. This procedure constitutes "the micro model." A macro approach employs the linear regression for estimating the unit rent per day $\cdot$ m^2, where the factors describing the features of apartment buildings together with 0-1 dummy variables for each subgroup are used as independent variables. For estimating the regression coefficients, the cross validation method is used. More formally, the whole data set is decomposed into 5 groups of equal size randomly. Four groups are used to estimate the regression coefficients, and the results would be applied to the remaining fifth group so as to test the accuracy of the regression model. This process is repeated 5 times for every possible combination of the 4 groups. The ultimate estimated values of the regression coefficients are determined by choosing those which achieve the minimum sum of the squared relative errors between the real data and the estimated values based on the regression model within the testing fifth group. This approach is called "the macro model."

Generally, the estimated values of the micro model and the macro model would not coincide. The final estimated value is obtained through a linear combination of the estimated results of the two models. More specifically, given the combination coefficient α, the new estimated value is determined as the sum of α times the result of the micro model and $(1-\alpha)$ times that of the macro model. The best final estimated value is determined by finding α^* so as to minimize the sum of the squared relative errors between the real data and the combined estimated values.

The structure of this paper is as follows. In Section 2, we introduce the data used for the analysis. Decomposition of the data set is described in Section 3. In Section 4, equations for deriving the unit rents and the unit opportunity costs are established. Procedures of evaluating the estimated unit rents through the micro model and the macro model are analyzed in Sections 5 and 6 respectively. Derivation of the final estimated value given by the combined micro-macro approach is discussed in Section 7, where the linear combination coefficient is also obtained explicitly. Section 8 presents a number of numerical results, demonstrating speed and accuracy of the algorithm proposed in this paper. Some concluding remarks are given in Section 9.

2 Data Description

The set of data to be employed in this study consists of revenue and other related information, as of December 2009, about 435 apartment buildings within the 23 wards in metropolitan Tokyo, which collectively constitute a variety of real estate funds listed in J-REIT (Japan Real Estate Investment Trust) between 2001 through 2009. Figure 1 exhibits the locations of these apartment buildings.

The attributes of each apartment building, necessary for the study, are classified into three categories: 1) those attributes with values that are invariant over time; 2) those attributes with values that change over time; 3) those attributes that describe certain performance-related indices. These attributes are listed in Tables 1 through 3 respectively. The suffix "ac" indicates that the underlying variable represents the value at the time of acquirement of the apartment building, while the suffix "pre" means the value at the present time. Throughout the paper, the present time τ^{pre} is assumed to be at the end of December 31, 2009. We note that any apartment building may be included in only one real estate fund at any given time. Furthermore, a record ID is assigned to each pair of an apartment building and the associated real estate fund. For example, if an apartment building belonged to a real estate fund, this fund was terminated, and then the same apartment building was included to another real estate fund later, such a case would be treated by assigning two separate record IDs to physically the same apartment building.

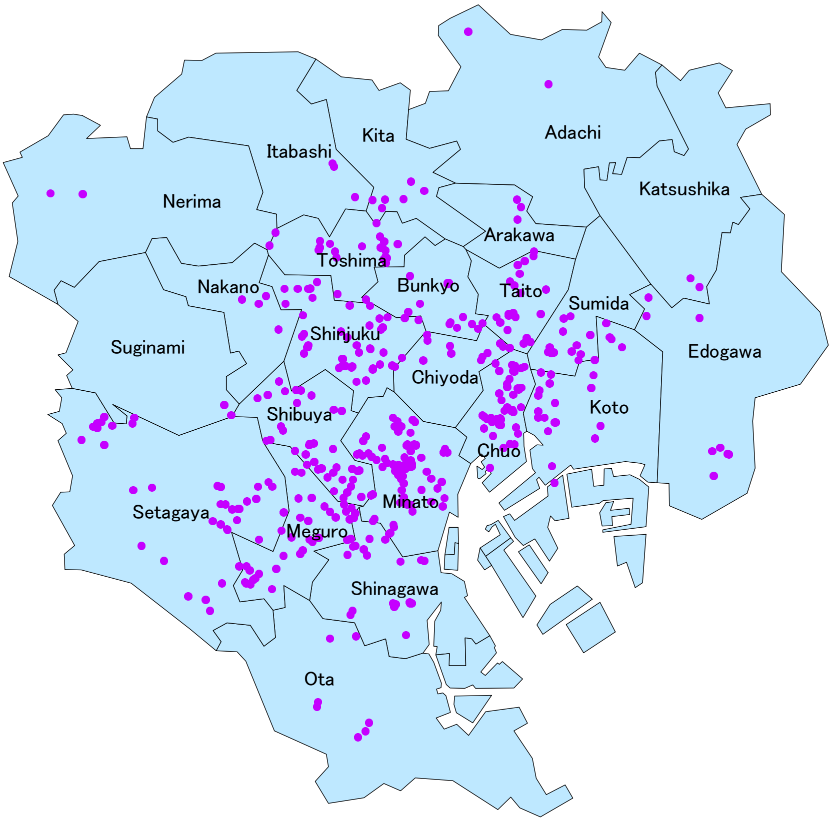

Figure 1: Locations of Apartment Buildings under Study

Table 1: Attributes Invariant over Time

τ_i^{ac}	the point in time at which Apartment Building i was included in a real estate fund under consideration
P_i^{ac}	the acquired price of Apartment Building i at τ_i^{ac}
FA_i^{ac}	the total floor area of Apartment Building i at τ_i^{ac}
UP_i^{ac}	the unit acquired price of Apartment Building i given by P_i^{ac}/FA_i^{ac}
BT_i	the structural type of Apartment Building $i \in \{S, RC, SRC\}$ †
ST_i	the nearest rail station of Apartment Building i
DS_i	the walking distance from Apartment Building i to ST_i in minutes

† : $S =$ Steal Structure, $RC =$Reinforced Concrete Structure, $SRC =$Steal Reinforced Concrete Structure

Table 2: Attributes Changing over Time

$RA_i(t)$	the total rented floor area of Apartment Building i at time t in m^2
$RE_i(t)$	the total rent revenue of Apartment Building i reported in the previous fiscal year before time t in Japanese yen
$C_i(t)$	the total expense of Apartment Building i reported in the previous fiscal year before time t in Japanese yen
$D_i(t)$	the number of days in which Apartment Building i had at least one tenant in the previous fiscal year before time t
$URE_i(t)$	the unit rent revenue of Apartment Building i per day $\cdot$ m^2 given by $RE_i(t)/\{RA_i(t) \times D_i(t)\}$
$UC_i(t)$	the unit cost of Apartment Building i per day $\cdot$ m^2 given by $C_i(t)/\{RA_i(t) \times D_i(t)\}$

Table 3: Performance-Related Indices

r_i	the expected exceeded return of Apartment Building i in percent of the total cost
$OPP_i(r_i,t)$	the opportunity cost per day $\cdot$ m^2 of Apartment Building i at time t, given r_i

3 Decomposition of Metropolitan Tokyo into Three Regions

In this paper, the macro model would be analyzed based on the linear regression approach. Hence, it would be wise to decompose the data set into subgroups so as to reduce the variance of UP_i^{ac} in Table 1, while keeping the size of each subgroup reasonably large. The linear regression can then be conducted within each subgroup separately. For this purpose, we consider three geographical regions and four intervals for UP_i^{ac} with total of $3 \times 4 = 12$ subgroups. Table 4 describes the four intervals for UP_i^{ac}, while Table 5 exhibits the three geographical regions. In Table 6, the distribution of the apartment buildings under consideration is shown over the twelve subgroups. We note that Region A is biased toward more valuable assets, while Region C contains only assets in the two lower intervals. Region B is between Region A and Region C. Figure 2 redraws Figure 1, where the locations of apartment buildings under study are depicted with distinction of the three geographical regions and the four intervals for UP_i^{ac}. The mean and the standard deviation of UP_i^{ac} are given for each subgroup in Table 7. We note that the standard deviations of UP_i^{ac} over 9 subgroups are reduced substantially by a factor of 0.33 or more in comparison with that for the entire apartment records, as expected.

4 Unit Rent Revenue $URE_i(t)$ and Unit Opportunity Cost $OPP_i(r_i,t)$

The purpose of this section is to establish a computational procedure for finding the Unit Rent Revenue $URE_i(t)$ and the Unit Cost $UC_i(t)$ as well as the Unit Opportunity Cost $OPP_i(r_i,t)$ of Apartment Building i at time t based on the real data, where r_i is the expected exceeded return of Apartment Building i in percent of the total cost given in Table 3. In this approach, only local information around Apartment Building i would be used.

Since the financial information associated with Apartment Building i would be provided only at

Table 4: Categories of Apartment Buildings by Unit Acquired Price UP^{ac} (in ¥10^3)

I			UP^{ac}	$\leq$	700
II	700	$<$	UP^{ac}	$\leq$	900
III	900	$<$	UP^{ac}	$\leq$	1,300
IV	1,300	$<$	UP^{ac}		

Table 5: Categories of Apartment Buildings by Region

Region A	Minato, Shibuya, Shinjuku, Chiyoda, Shinagawa, Meguro
Region B	Setagaya, Koto, Chuo, Bunkyo, Ota, Sumida, Toshima
Region C	Katsushika, Edogawa, Arakawa, Suginami, Adachi, Itabashi, Kita, Nakano, Taito, Nerima

Table 6: Number of Apartment Buildings in Each Subgroup

	Region A	Region B	Region C	Total
I	21	60	38	119
II	75	79	18	172
III	118	16	0	134
IV	10	0	0	10
Total	224	155	56	435

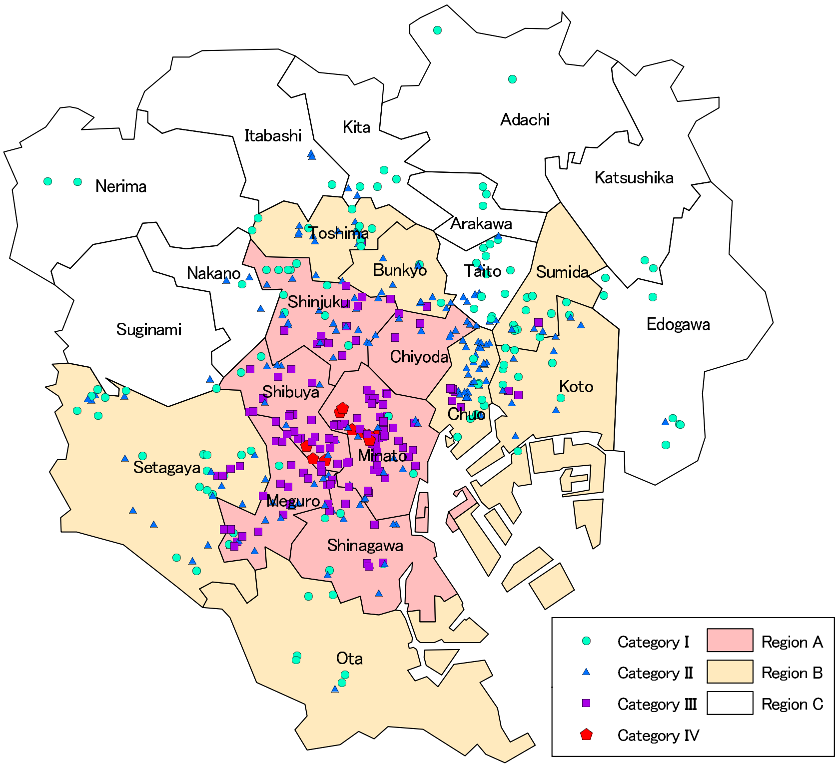

Figure 2: Categories of Apartment Buildings

Table 7: Range, Mean and Standard Deviation of Unit Acquired Prices of Apartment Buildings for Each Region

		Region *A*	Region *B*	Region *C*
I	#	21	60	38
	Range (in ¥10^3)	82.0 – 694.7	411.5 – 698.1	196.9 – 699.2
	Mean (in ¥10^3)	590.2	621.5	533.0
	STD	55.2	67.9	93.4
II	#	75	79	18
	Range (in ¥10^3)	700.5 – 896.3	703.2 – 898.1	701.3 – 853.9
	Mean (in ¥10^3)	823.7	788.4	763.1
	STD	55.5	53.5	45.4
III	#	118	16	0
	Range (in ¥10^3)	900.1 – 1288.3	901.2 – 1162.9	-
	Mean (in ¥10^3)	1033.0	998.1	-
	STD	99.8	76.4	-
IV	#	10	0	0
	Range (in ¥10^3)	1316.6 – 1912.3	-	-
	Mean (in ¥10^3)	1499.1	-	-
	STD	170.4	-	-
Total	#	224	155	56
	Range (in ¥10^3)=82.0 – 1912.3, Mean (in ¥10^3)=829.0, STD=216.6			

the end of each fiscal year for accounting the real estate fund containing Apartment Building i, it is necessary to reevaluate the latest information available at the present time τ^{pre}. The consumer price index, which is available as daily data from Ministry of Internal Affairs and Communications of the Japanese Government (2010), would be employed for this purpose. More specifically, let t_i be the nearest end point of a fiscal year before τ^{pre} for Apartment Building i. Then the unit rent revenue and the unit cost for Apartment Building i reevaluated at τ^{pre} can be obtained as

$$URE_i(\tau^{pre}) = URE_i(t_i) \cdot \frac{CPI(\tau^{pre})}{CPI(t_i)} \ ; \quad UC_i(\tau^{pre}) = UC_i(t_i) \cdot \frac{CPI(\tau^{pre})}{CPI(t_i)} \ , \tag{4.1}$$

where $CPI(t)$ denotes the consumer price index for day t.

The value of $URE_i(\tau^{pre})$ represents the income side of Apartment Building i, while $UC_i(\tau^{pre})$ describes a portion of the cost side of Apartment Building i. The additional cost factor to be considered is the opportunity cost $OPP_i(r_i,t)$, given by

$$OPP_i(r_i, \tau^{pre}) = URE_i(\tau^{pre}) \cdot \frac{1}{1+r_i} - UC_i(\tau^{pre}) \ . \tag{4.2}$$

The economic interpretation of (4.2) may be easily seen by rewriting it as

$$URE_i(\tau^{pre}) = \Big\{ UC_i(\tau^{pre}) + OPP_i(r_i, \tau^{pre}) \Big\} \times (1+r_i) \ , \tag{4.3}$$

i.e. the revenue should be equal to the total cost consisting of the actual cost UC_i and the opportunity cost OPP_i times $1+r_i$ where r_i is the expected exceeded return of Apartment Building i.

In summary, we obtain both $URE_i(t_i)$ and $UC_i(t_i)$ from the data, which in turn yield $URE_i(\tau^{pre})$ and $UC_i(\tau^{pre})$ from (4.1). Assuming that r_i is known, one can then find $OPP_i(r_i, \tau^{pre})$ from (4.2).

5 Alternative Evaluation of Unit Opportunity Cost $OPP_i(\tau^{pre})$ Based on Value of the Nearest Rail Station for Estimating Expected Exceeded Returns and Unit Rent Revenue $URE_i(\tau^{pre})$: Micro Model

In the previous section, the Unit Rent Revenue $URE_i(\tau^{pre})$ and the Unit Opportunity Cost $OPP_i(\tau^{pre})$ of Apartment Building i are obtained based on the real data. The ultimate purpose of this paper, however, is to establish an algorithmic framework to estimate an appropriate unit rent revenue of an apartment building without any past financial information. Toward this goal, in this section, we introduce the concept of the value of a rail station as an intermediary step. The value of a rail station would be assessed using the unit opportunity costs of the apartment buildings having that station as their nearest rail station. Once this value is determined from the real data, it can be used to estimate an appropriate unit rent revenue of any apartment building near the station.

Let $N(s)$ be the set of apartment buildings having Station s as their nearest rail station, that is,

$$N(s) = \{ i \mid ST_i = s \} \ . \tag{5.1}$$

The mean and the standard deviation of the unit opportunity costs of the apartment buildings in $N(s)$ at time τ^{pre} are denoted by $\mu_{OPP}(s, \tau^{pre})$ and $\sigma_{OPP}(s, \tau^{pre})$ respectively. More specifically, one has

$$\mu_{OPP}(s, \tau^{pre}) = \frac{1}{|N(s)|} \sum_{i \in N(s)} OPP_i(r_i, \tau^{pre}) \tag{5.2}$$

and

$$\sigma_{OPP}(s, \tau^{pre}) = \sqrt{\frac{\sum_{i \in N(s)} \{ OPP_i(r_i, \tau^{pre}) - \mu_{OPP}(s, \tau^{pre}) \}^2}{|N(s)| - 1}} \ , \tag{5.3}$$

where $|A|$ denotes the cardinality of a set A. We say that the pair $[\mu_{OPP}(s,\tau^{pre}),\ \sigma_{OPP}(s,\tau^{pre})]$ determines the value of Station s.

Suppose Apartment Building i has Station s as the nearest rail station, i.e. $ST_i = s$. We assume that the opportunity cost of Apartment Building i can be evaluated alternatively solely based on the distance between Apartment Building i and Station s, denoted by DS_i, together with $[\mu_{OPP}(s,\tau^{pre}),\ \sigma_{OPP}(s,\tau^{pre})]$, where $\mu_{OPP}(s,\tau^{pre})$ is adjusted by DS_i using $\sigma_{OPP}(s,\tau^{pre})$ as the adjustment unit. In order to distinguish this alternative derivation from (4.2), we write $OPP_i^{\#}(\tau^{pre})$ in place of $OPP_i(r_i,\tau^{pre})$. More formally, let $\mu_{DS}(s)$ be defined by

$$\mu_{DS}(s) = \frac{1}{|N(s)|}\sum_{i\in N(s)} DS_i \ . \tag{5.4}$$

It is assumed that the value of $OPP_i^{\#}(\tau^{pre})$ can be given by

$$OPP_i^{\#}(\tau^{pre}) = \mu_{OPP}(s,\tau^{pre}) + \left\{1 - \frac{DS_i}{\mu_{DS}(s)}\right\}\sigma_{OPP}(s,\tau^{pre}) \ . \tag{5.5}$$

We note that $OPP_i^{\#}(\tau^{pre})$ is greater than $\mu_{OPP}(s,\tau^{pre})$ if and only if DS_i is smaller than $\mu_{DS}(s)$. Given r_i, one sees in parallel with (4.3) that

$$URE_i^{\#}(r_i,\tau^{pre}) = \left\{UC_i(\tau^{pre}) + OPP_i^{\#}(\tau^{pre})\right\} \times (1+r_i) \ , \tag{5.6}$$

where the symbol # is used in the same manner as for $OPP_i^{\#}(\tau^{pre})$.

In general, $URE_i(\tau^{pre})$ obtained from the data as in (4.1) and $URE_i^{\#}(r_i,\tau^{pre})$ given in (5.6) are not equal. This discrepancy, however, enables one to estimate the expected exceeded return by setting its value so as to minimize the total discrepancy among apartment buildings within each subgroup having the same nearest rail station, provided that the same value of the expected exceeded return would prevail within such apartment buildings. For this purpose, let this group of apartment buildings be denoted by

$$Bu(k,z,s) = \{i \mid I_i = k, Z_i = z, ST_i = s\} \ ,$$

where I_i, Z_i, and ST_i denotes categories of Apartment Building i by unit acquired price UP^{ac}, categories of Apartment Building i by region, and nearest rail station of Apartment Building i respectively. It is assumed that

$$r_i = r_j = r(k,z,s) \ \text{ for all } \ i,j \in Bu(k,z,s) \ .$$

Then the estimated value of the expected exceeded return for $Bu(k,z,s)$ can be obtained as

$$r^*(k,z,s) = \underset{r(k,z,s)\geq 0}{\operatorname{argmin}} \left\{ \sqrt{\sum_{i\in Bu(k,z,s)} \left\{ \frac{URE_i^{\#}(r(k,z,s),\tau^{pre}) - URE_i(\tau^{pre})}{URE_i(\tau^{pre})} \right\}^2} \right\} \ . \tag{5.7}$$

Once $r^*(k,z,s)$ is determined, the value of Station s can be updated by repeating (4.2), (5.2) and (5.3), where r_i is replaced by $r^*(k,z,s)$. More specifically, for $i \in Bu(k,z,s)$, we define

$$OPP_i^*(\tau^{pre}) = URE_i(\tau^{pre}) \cdot \frac{1}{1+r^*(k,z,s)} - UC_i(\tau^{pre}) \ , \tag{5.8}$$

$$\mu_{OPP}^*(s,\tau^{pre}) = \frac{1}{|N(s)|}\sum_{i\in N(s)} OPP_i^*(\tau^{pre}) \ , \tag{5.9}$$

and

$$\sigma^*_{OPP}(s,\tau^{pre}) = \sqrt{\frac{\sum_{i\in N(s)}\left(OPP^*_i(\tau^{pre}) - \mu^*_{OPP}(s,\tau^{pre})\right)^2}{|N(s)|-1}}\ . \tag{5.10}$$

The above procedure establishes a micro approach for estimating the unit rent revenue of Apartment Building n for $n \in Bu(k,z,s)$ at time τ^{pre}, for which only the unit cost at time τ^{pre} and the distance to the nearest rail station s are known. In other words, we are in a position to compute $URE^{\#*}_n(\tau^{pre})$ through the following steps.

$$OPP^{\#*}_n(\tau^{pre}) = \mu^*_{OPP}(s,\tau^{pre}) + \left\{1 - \frac{DS_n}{\mu_{DS}(s,\tau^{pre})}\right\}\sigma^*_{OPP}(s,\tau^{pre}) \tag{5.11}$$

$$URE^{\#*}_n(\tau^{pre}) = \left\{UC_n(\tau^{pre}) + OPP^{\#*}_n(\tau^{pre})\right\} \times \{1 + r^*(k,z,s)\} \tag{5.12}$$

We say that a sequence of the above procedures would constitute "the micro model." Equation (5.11) would be used to feed one variable to the regression equation in the macro model, as we will see.

6 Linear Regression for Estimating Unit Rent Revenue $URE_i(\tau^{pre})$: Macro Model

The purpose of this section is to establish a linear regression model for estimating the Unit Rent Revenue $URE_i(\tau^{pre})$. We first consider 4 independent variables listed in Table 8, where their values are adjusted to the present value at $t = \tau^{pre}$ using CPI, as in (4.1).

Table 8: Candidate Independent Variables

UC_i	the unit cost of Apartment Building i per day $\cdot$ m^2
$OPP^{\#*}_i$	estimated value of Apartment Building i per day $\cdot$ m^2 via the micro model
Y_i	the age of Apartment Building i
WT_i	the walking time from Apartment Building i to ST_i in minutes

Let $SG(k,z)$ be the set of apartment buildings in the subgroup (k,z), i.e. $SG(k,z) = \{i \mid I_i = k, Z_i = z\}$, where $k \in \mathbf{I} = \{\mathrm{I},\mathrm{II},\mathrm{III},\mathrm{IV}\}$ and $z \in \mathbf{Z} = \{A,B,C\}$. It may be desirable to establish a separate linear regression equation for each subgroup $(k,z) \in \mathbf{I}\times\mathbf{Z}$. However, the sample sizes are not sufficiently large for some subgroups. In order to overcome this difficulty, the following dummy variables are introduced. Since any subgroup in $\mathbf{Null} = \{(\mathrm{III},C),(\mathrm{IV},B),(\mathrm{IV},C)\}$ has no data, we define $\mathbf{SG} = (\mathbf{I}\times\mathbf{Z}) \setminus (\mathbf{Null} \cup \{(\mathrm{I},A)\})$. Given Apartment Building i, we then define, for each subgroup $(k,z) \in \mathbf{SG}$,

$$Dummy_{(k,z),i} = \begin{cases} 1 & \text{if } i \in SG(k,z) \\ 0 & \text{else} \end{cases}\ . \tag{6.1}$$

Here, (I,A) is eliminated from $\mathbf{SG}$ because an apartment building with all the dummy variables being 0 is defined to belong to (I,A). Concerning the structural type of Apartment Building i, denoted by BT_i as in Table 1, we also define the following two dummy variables.

$$Dummy_{S,i} = \begin{cases} 1 & \text{if } BT_i = S \\ 0 & \text{else} \end{cases}\ ; \qquad Dummy_{SRC,i} = \begin{cases} 1 & \text{if } BT_i = SRC \\ 0 & \text{else} \end{cases}\ . \tag{6.2}$$

By evaluating AIC (Akaike Information Criterion, see e.g. Burnham & Anderson (2002)) for every possible combination of the independent variables using all the data, the set of independent variables achieving the minimum is selected. The following linear regression equation resulted from this procedure.

$$\begin{aligned} URE_i(\tau^{pre}) \quad = \quad & \beta_0 + \beta_{UC} UC_i(\tau^{pre}) + \beta_{OPP} OPP_i^{\#*}(\tau^{pre}) + \beta_{WT} WT_i \\ & + \beta_S Dummy_{S,i} + \beta_{SRC} Dummy_{SRC,i} + \sum_{(k,z)\in\mathbf{SG}} \beta_{(k,z)} Dummy_{(k,z),i} + \varepsilon_i \end{aligned} \tag{6.3}$$

For estimating the regression coefficients in (6.3), we employ the cross validation approach. Namely, the whole data set is decomposed into 5 groups of equal size randomly. Four groups are used to estimate the regression coefficients, and the results would be applied to the remaining fifth group so as to test the accuracy of the regression model. This process is repeated 5 times for every possible combination of the 4 groups. The ultimate estimated values of the regression coefficients are determined by choosing those which achieve the minimum sum of the squared relative errors between the real data and the estimated values based on the regression model within the testing fifth group. Consequently, β_0^{**}, β_{UC}^{**}, β_{OPP}^{**}, β_{WT}^{**},β_S^{**}, β_{SRC}^{**}, and $\beta_{(k,z)}^{**}$ for $(k,z)\in\mathbf{SG}$ are obtained as summarized in Table 9.

Table 9: Estimated Regression Coefficients

Coefficients	Estimate	Std.Error	t-Value	$Pr(>\|x\|)$	Significance Level
β_0^{**}	81.07052	5.40303	15.005	2.00e-16	⋆⋆⋆
β_{UC}^{**}	0.66837	0.04636	14.416	2.00e-16	⋆⋆⋆
β_{OPP}^{**}	0.22782	0.04167	5.468	7.82e-08	⋆⋆⋆
β_{WT}^{**}	-0.84045	0.24833	-3.384	0.00078	⋆⋆⋆
β_S^{**}	-38.30547	6.87045	-5.575	4.42e-08	⋆⋆⋆
β_{SRC}^{**}	-4.69630	1.93340	-2.429	0.01556	⋆
$\beta_{(\mathrm{II},A)}^{**}$	19.64121	3.99299	4.919	1.25e-06	⋆⋆⋆
$\beta_{(\mathrm{III},A)}^{**}$	34.09839	3.93672	8.662	2.00e-16	⋆⋆⋆
$\beta_{(\mathrm{IV},A)}^{**}$	97.13690	6.51594	14.908	2.00e-16	⋆⋆⋆
$\beta_{(\mathrm{I},B)}^{**}$	0.20764	4.07643	0.051	0.95940	
$\beta_{(\mathrm{II},B)}^{**}$	12.33288	3.96643	3.109	0.00200	⋆⋆
$\beta_{(\mathrm{III},B)}^{**}$	27.76009	5.36472	5.175	3.54e-07	⋆⋆⋆
$\beta_{(\mathrm{I},C)}^{**}$	-4.13632	4.40108	-0.940	0.34784	
$\beta_{(\mathrm{II},C)}^{**}$	12.83417	5.17115	2.482	0.01346	⋆
Significance Level: ⋆⋆⋆ 0.001, ⋆⋆ 0.01, ⋆ 0.05 Adjusted $R^2 = 0.761$					

It should be noted that all the coefficients except $\beta_{(\mathrm{I},B)}^{**}$ and $\beta_{(\mathrm{I},C)}^{**}$ satisfy the 5% significance level with adjusted $R^2 = 0.761$ which is fairy high. Both β_{UC}^{**} and β_{OPP}^{**} are positive, but β_{WT}^{**} is negative. This means that the unit rent revenue increases as a function of the unit cost or the unit opportunity cost, while it decreases as the walking distance from the nearest rail station increases, showing the consistency with our common sense. As for the dummy variables concerning the structural type of apartment buildings, one has $\beta_S^{**} < \beta_{SRC}^{**} < 0$ while it may be expected to observe $\beta_S^{**} < 0 < \beta_{SRC}^{**}$ since SRC is supposedly stronger than RC which is the standard of our regression model. There are 337 apartment buildings of RC type while only 92 apartment buildings of SRC type, and this imbalance might have contributed to the reversed order. The estimated coefficients for the dummy variables for $(k,z) \in \mathbf{SG}$ agree with our intuition. For example, $0 < \beta_{(\mathrm{II},A)}^{**} < \beta_{(\mathrm{III},A)}^{**} < \beta_{(\mathrm{IV},A)}^{**}$, $0 < \beta_{(\mathrm{I},B)}^{**} < \beta_{(\mathrm{II},B)}^{**} < \beta_{(\mathrm{III},B)}^{**}$, $0 < \beta_{(\mathrm{II},B)}^{**} < \beta_{(\mathrm{II},A)}^{**}$, $0 < \beta_{(\mathrm{III},B)}^{**} < \beta_{(\mathrm{III},A)}^{**}$, $\beta_{(\mathrm{I},C)}^{**} < 0$, and $\beta_{(\mathrm{I},C)}^{**} < \beta_{(\mathrm{II},C)}^{**}$.

For Apartment Building n with $OPP_n^{\#*}$ estimated through the micro model, its unit rent revenue can now be estimated by

$$\begin{aligned} URE_n^{**}(\tau^{pre}) &= \beta_0^{**} + \beta_{UC}^{**} UC_n(\tau^{pre}) + \beta_{OPP}^{**} OPP_n^{\#*}(\tau^{pre}) + \beta_{WT}^{**} WT_n \\ &\quad + \beta_S^{**} Dummy_{S,n} + \beta_{SRC}^{**} Dummy_{SRC,n} + \sum_{(k,z)\in\mathbf{SG}} \beta_{(k,z)}^{**} Dummy_{(k,z),n} \, . \end{aligned} \tag{6.4}$$

As before, we say that a sequence of the above procedures would constitute "the macro model."

7 Combined Micro-Macro Approach

We have seen that the unit rent revenue can first be estimated by the micro model as $URE_n^{\#*}(\tau^{pre})$ in (5.12) , and then separately by the macro model as $URE_n^{**}(\tau^{pre})$ in (6.4). It is then natural to consider a linear combination of the two estimated values, i.e.

$$\widehat{URE}_n(\alpha, \tau^{pre}) = \alpha \times URE_n^{\#*}(\tau^{pre}) + (1-\alpha) \times URE_n^{**}(\tau^{pre}) \, . \tag{7.1}$$

In order to determine the best final estimated value, we set α so as to minimize the sum of the squared relative errors between the real data and the combined estimated values. More specifically, let L be the set of all the apartment buildings in the data set. The optimal value α^* is then determined by

$$\alpha^* = \underset{0\le\alpha\le1}{\operatorname{argmin}} \left\{ \sum_{i\in L} \left(\frac{\widehat{URE}_i(\alpha,\tau^{pre}) - URE_i(\tau^{pre})}{URE_i(\tau^{pre})} \right)^2 \right\} \, . \tag{7.2}$$

After a little algebra, one finds that the derivative of the sum in (7.2) with respect to α achieves the value 0 at $\widehat{\alpha}$ given by

$$\widehat{\alpha} = \frac{\sum_{i\in L} \frac{\left(URE_i(\tau^{pre}) - URE_i^{**}(\tau^{pre})\right)\left(URE_i^{\#*}(\tau^{pre}) - URE_i^{**}(\tau^{pre})\right)}{URE_i(\tau^{pre})^2}}{\sum_{i\in L} \left(\frac{URE_i^{\#*}(\tau^{pre}) - URE_i^{**}(\tau^{pre})}{URE_i(\tau^{pre})} \right)^2} \, . \tag{7.3}$$

The optimal α^* can then be specified as

$$\alpha^* = \begin{cases} 0 & \text{if } \widehat{\alpha} \le 0 \\ \widehat{\alpha} & \text{if } 0 < \widehat{\alpha} < 1 \\ 1 & \text{if } \widehat{\alpha} \ge 1 \end{cases} \, . \tag{7.4}$$

For the data set described in Sections 2 and 3, one has

$$\alpha^* = 0.760 \, . \tag{7.5}$$

With this α^*, one obtains the ultimate estimated value as

$$\widehat{URE}_n(\alpha^*, \tau^{pre}) = \alpha^* \times URE_n^{\#*}(\tau^{pre}) + (1-\alpha^*) \times URE_n^{**}(\tau^{pre}) \, . \tag{7.6}$$

We note from (7.5) and (7.6) that the combined micro-macro approach for the data set weighs the micro approach more than the macro approach approximately by a factor of 3. The whole procedure of the combined micro-macro approach is summarized in Figure 3 below.

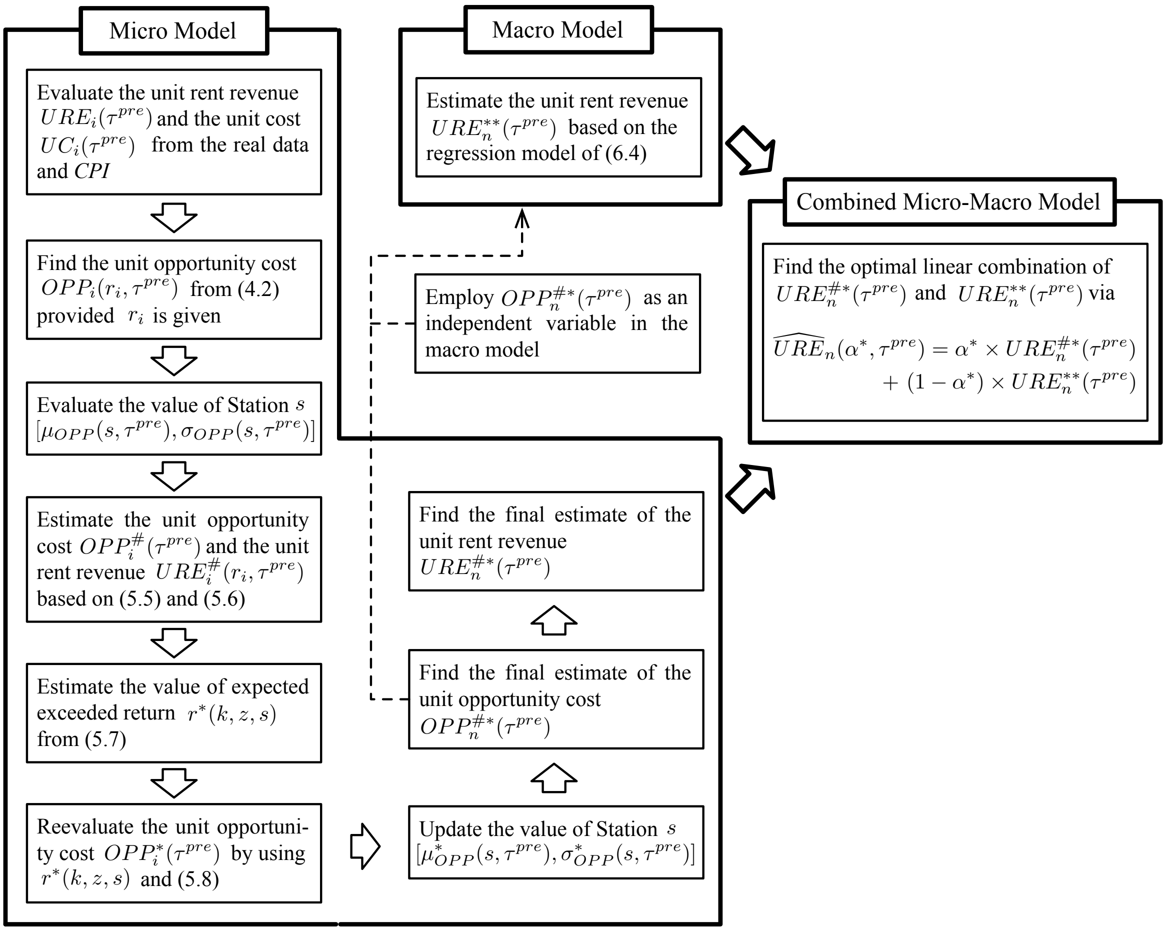

Figure 3: Combined Micro-Macro Approach for Estimating Unit Rent Revenue

8 Numerical Results

In this section, numerical results are presented for demonstrating speed and accuracy of the estimation procedure for Unit Rent Revenue(URE) proposed in this paper. In order to evaluate the accuracy of the micro model, we define

$$ACC_{mic}(K) = \sqrt{\frac{1}{|K|}\sum_{i\in K}\Big(\frac{URE_i^{\#*}(\tau^{pre}) - URE_i(\tau^{pre})}{URE_i(\tau^{pre})}\Big)^2}\,, \tag{8.1}$$

where K is an arbitrary subset of the apartment buildings in the data set L. Similarly, the accuracy of the macro model $ACC_{mac}(K)$ and that of the combined micro-macro model $ACC_{mic-mac}(K)$ can be defined as

$$ACC_{mac}(K) = \sqrt{\frac{1}{|K|}\sum_{i\in K}\Big(\frac{URE_i^{**}(\tau^{pre}) - URE_i(\tau^{pre})}{URE_i(\tau^{pre})}\Big)^2}\,, \tag{8.2}$$

and

$$ACC_{mic-mac}(K) = \sqrt{\frac{1}{|K|}\sum_{i\in K}\Big(\frac{\widehat{URE}_i(\alpha^*, \tau^{pre}) - URE_i(\tau^{pre})}{URE_i(\tau^{pre})}\Big)^2}\,. \tag{8.3}$$

In Table 10, the accuracy measures for the three estimation algorithms are provided for each subgroup $(k,z) \in (\mathbf{I}\times\mathbf{Z})\setminus\mathbf{Null}$. Table 11 exhibits the three accuracy measures for individual stations, each of which has more than 6 apartment buildings within its vicinity. From these two tables together with Table 7, the following observations can be made.

1. For the entire data set $K = L$, the combined micro-macro approach is superior with accuracy of $ACC_{mic-mac}(L) = 0.080$ to both the micro approach having the accuracy of $ACC_{mic}(L) = 0.083$ and the macro approach with $ACC_{mac}(L) = 0.106$.

2. The combined micro-macro approach is comparable with the micro approach and they jointly outperform the macro approach when the standard deviation of the unit acquired price is small, say less than 60 corresponding to K =(II, C), (II, B), (I, A), or (II, A). Also when the sample size is small, say less than 60 for K = (I, B), (III, B), (I, C) or (IV, A).

3. The macro approach seems to be competitive against the other two approaches only when both the standard deviation and sample size are large, as can be seen at K = (III, A) with the standard deviation of 99.8 and the sample size of 118.

4. For the region A, the accuracy for the combined micro-macro approach as well as the micro approach decreases as the price category increases. This trend is reversed for the regions B and C. Perhaps, this is so because the standard deviation of the unit acquired price increases from I to IV in A, while the standard deviation decreases or is comparable as the price category changes from I to IV in B and C.

5. In Table 11, the three accuracies are exhibited for individual station with $K = N(s)$ for Station s satisfying $|N(s)| \geq 7$. The combined micro-macro approach is the best for 6 stations out of 12 stations, followed by the macro approach outperforming the other two approaches for 5 stations. The micro approach is least competitive against the other two approaches, achieving the best performance only for 2 stations. Further study is needed so as to identify the characteristics of a station for which one approach outperforms the other two approaches.

Table 10: Accuracy of Micro Model, Macro Model and Combined Micro-Macro Model for Individual Subgroups

	A			*B*			*C*			All		
	mic	*mac*	*mic-mac*	*mic*	*mac*	*mic-mac*	*mic*	*mac*	*mic-mac*	*mic*	*mac*	*mic-mac*
#	21			60			38			119		
I	**0.034**	0.088	0.041	0.077	0.109	**0.076**	**0.066**	0.162	0.080	**0.068**	0.125	0.072
#	75			79			18			172		
II	0.068	0.097	**0.067**	0.079	0.084	**0.073**	**0.020**	0.107	0.035	0.070	0.092	**0.067**
#	118			16			0			134		
III	0.106	**0.096**	0.097	0.057	0.098	**0.056**	-	-	-	0.101	0.096	**0.093**
#	10			0			0			10		
IV	0.146	0.188	**0.141**	-	-	-	-	-	-	0.146	0.188	**0.141**
#	224			155			56			435		
All	0.092	0.101	**0.087**	0.077	0.096	**0.073**	**0.056**	0.146	0.069	0.083	0.106	**0.080**

Table 11: Accuracy of Micro Model, Macro Model and Combined Micro-Macro Model for Individual Stations

Nearest Rail Stations	#	ACC_{mic}	ACC_{mac}	$ACC_{mic-mac}$	Number of Apartment Buildings in the Subgroup
Azabu-juban	19	0.169	**0.131**	0.151	(III,A):19
Ebisu	10	0.090	0.085	**0.082**	(II,A):6, (III,A):4
Toritsu-daigaku	9	0.066	0.069	**0.058**	(I,A):2, (II,A):2, (III,A):4, (III,B):1
Gakugei-daigaku	9	**0.073**	0.093	**0.073**	(II,A):5, (III,A):3, (II,B):1
Shibuya	9	0.171	**0.140**	0.162	(III,A):9
Shintomicho	8	0.047	0.065	**0.046**	(II,B):7, (III,B):1
Otsuka	8	0.058	0.065	**0.055**	(I,B):4, (II,B):3, (III,B):1
Waseda	7	**0.035**	0.080	0.038	(II,A):2, (III,A):4, (II,B):1
Oimachi	7	0.055	**0.043**	0.047	(I,A):1, (II,A):3, (III,A):3
Hatchobori	7	0.106	**0.075**	0.093	(I,B):3, (II,B):4
Roppongi-itchome	7	0.133	**0.104**	0.114	(I,A):1, (II,A):2, (III,A):4
Hiro-o	7	0.122	0.136	**0.119**	(II,A):3, (III,A):3, (IV,A):1

9 Concluding Remarks

In this paper, a new comprehensive scheme has been developed for estimating rental prices of large apartment buildings based on a set of real data in the metropolitan Tokyo. The data set is first decomposed into subgroups along two axes so as to reduce the variance of the unit acquired price of the apartment buildings in each subgroup. The first axis is the unit acquired price itself, decomposing the price range into 4 intervals. The second axis is the regional characteristics of the 23 wards in the metropolitan Tokyo, grouping them into 3 geographical regions: one with concentration of expensive large apartment buildings, another having large apartment buildings of low unit acquired prices as a majority, and the third between the two. The combined micro-macro approach would be implemented separately in each of $4 \times 3 = 12$ subgroups. The micro model is built upon a new concept of "the value of a rail station" expressed in terms of the mean and the variance of the unit rents per day $\cdot$ m^2 of the large apartment buildings having the rail station as the nearest in common. The macro model employs linear regression using the features of apartment buildings as independent variables together with dummy variables corresponding to price-geographical ranges to which they belong. The final estimated value is then obtained by constructing the optimal linear combination of the results of the two models so as to minimize the sum of the squared relative errors between the real data and the estimated value.

For the entire data set, the combined micro-macro approach is superior to the micro approach or the macro approach alone. The combined micro-macro approach is comparable with the micro approach and they jointly outperform the macro approach when the standard deviation of the unit acquired price is small, say less than 60. However, the macro approach seems to be competitive against the other two approaches when both the standard deviation and sample size within a subgroup are large. It is necessary to establish a general guidance for deciding which approach to be employed under what conditions. This study is in progress and will be reported elsewhere.

References

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association*, 58(304), 933–942.

Boyle, M. A. & Kiel, K. A. (2001). A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities. *Journal of Real Estate Literature*, 9(2), 117–144.

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2009). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*. (doi: 10.1007/s11146-009-9177-z).

Burnham, K. P. & Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2nd edition.

Case, B. & Quigley, J. M. (1991). The Dynamics of Real Estate Prices. *The Review of Economics and Statistics*, 73(1), 50–58.

Conroy, S. J. & Milosch, J. L. (2009). An Estimation of the Coastal Premium for Residential Housing Prices in San Diego County. *The Journal of Real Estate Finance and Economics*. (doi: 10.1007/s11146-009-9195-x).

Debrezion, G., Pels, E., & Rietveld, P. (2007). The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis. *The Journal of Real Estate Finance and Economics*. (doi: 10.1007/s11146-007-9032-z).

Do, A. Q. & Grudnitski, G. (1995). Golf Courses and Residential House Prices: An Empirical Examination. *Journal of Real Estate Finance and Economics*, 10, 261–270.

Goodman, A. G. (1978). Hedonic Prices, Price Indices and Housing Markets. *Journal of Urban Economics*, 5, 471–484.

Hwang, M. & Quigley, J. M. (2009). Housing Price Dynamics in Time and Space: Predictability, Liquidity and Investor Returns. *The Journal of Real Estate Finance and Economics*. (doi: 10.1007/s11146-007-9207-x).

Jin, C., Huang, J., Sumita, U., & Lu, S. (2008). Development of Apartment Rent Estimation Model Based on an Integrated Macro-Micro Approach. *Journal of Real Estate Financial Engineering*, 3, 15–25. (in Japanese).

Ministry of Internal Affairs and Communications of the Japanese Government, Statistics Bureau. (2010).

Morancho, A. B. (2003). A hedonic valuation of urban green areas. *Landscape and Urban Planning*, 66, 35–41.

Poon, L. C. L. (1978). Railway Externalities and Residential Property Prices. *Land Economics*, 54(2), 218–227.

Quigley, J. M. (1995). A Simple Hybrid Model for Estimating Real Estate Price Indexes. *Journal of Housing Economics*, 4, 1–12.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Rosenthal, E. C. (2009). A Pricing Model for Residential Homes with Poisson Arrivals and a Sales Deadline. *The Journal of Real Estate Finance and Economics*. (doi: 10.1007/s11146-009-9191-1).

Sheppard, S. (1999). Hedonic Analysis of Housing Markets. In P. Cheshire & E. S. Mills (Eds.), *Handbook of Regional and Urban Economics*, volume 3 of *Applied Urban Economics* chapter 41, (pp. 1595–1635). Elsevier.

Tyrväinen, L. & Miettinen, A. (2000). Property Prices and Urban Forest Amenities. *Journal of Environmental Economics and Management*, 39, 205–223.