Hedonic Regression Models for Tokyo Condominium Sales

by Erwin Diewert

University of British Columbia

(Presentation by Chihiro Shimizu, Nihon University)

Hitotsubashi-RIETI International Workshop on Real Estate Market, Productivity, and Prices

October 13-14, 2016

Research Institute of Economy, Trade and Industry

1-3-1 Kasumigaseki Chiyoda-ku Tokyo, JAPAN

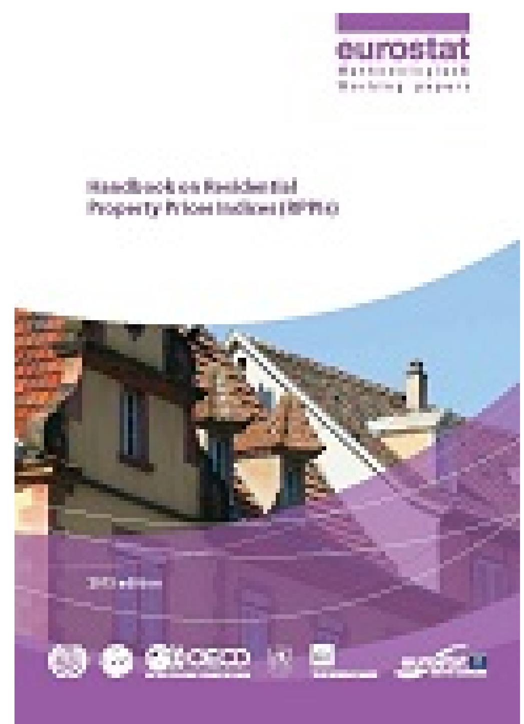
Background and Motivation (1)

- **Handbook on Residential Property Price Indexes from OECD, UN, IMF, BIS, World Bank from *EuroStat* (2013).**

- **Fenwick (2006):**
- As a *general macroeconomic indicator* (of inflation);
- As an input into the measurement of consumer price inflation; As an element in the calculation of

  household (real) wealth
- As a direct input into an analysis

  of *mortgage lender's exposure to risk of default*.



Shimizu, C., K. G. Nishimura and T. Watanabe (2010), "Residential Rents and Price Rigidity: Micro Structure and Macro Consequences," *Journal of Japanese and International Economy*, 24, 282-299.

Background and Motivation (2)

- **The International System of National Accounts asks countries to provide estimates for the value of assets held by the various sectors in the economy.**

- **These estimates are supposed to appear in the Balance Sheet Accounts of the country. An important asset for the Household Sector is the stock of housing.**

- **For many modeling purposes, it is important to not only have estimates for the value of the housing stock but to decompose the overall value into (additive) land and structure components and then to further decompose these value aggregates into constant quality price and quantity components.**

Background and Motivation (3)

- **This is not an easy task. When a housing property is sold, the selling price values the sum of the structure and land components and so a structure-land decomposition must be obtained by a modeling exercise.**

- **The problem of obtaining *constant quality* price components for the land and structure components of a housing unit is further complicated by the fact that housing units are almost always unique assets.**

→A dwelling unit is different from any other dwelling unit at the same period in time due to its *location*, which is unique (and as locations *vary for the same physical structure*, *the price of the land plot for the unit will generally change due to locational amenities*).

→The same dwelling unit compared over space will also be different due to *depreciation* and possible *renovations* to the structure.

Literature Reviews

- **In Asian countries, Shimizu, Nishimura and Watanabe (2010) compared Repeat Sales Indexes to Hedonic Indexes for Tokyo, while Deng, McMillen and Sing (2012) proposed a matching method for the Singapore housing market.**
- **Wu, Deng and Liu (2014) and Guo, Zheng, Geltner and Liu (2014) constructed price indexes for newly built houses in China.**

- **→There are very few papers that construct quality adjusted price indexes for condominium sales which is our focus, along with providing a method to decompose property sales into land and structure components.**

- **→Chapter 8 of Eurostat (2013), Diewert and Shimizu(2015) where a similar modeling strategy was applied to sales of detached dwellings. →Builder's Model**

Purposes

The paper fits a hedonic regression model to *the sales of condominium units* in Tokyo over the period 2000-2015.

Major Problems:

- **The selling price of a condo unit has two main characteristics: (i) the floor space area of the unit and (ii) the unit's share of the land area of the building. But how exactly can we decompose the total property value into these two components? And how can we determine the unit's share of land value?**
- **Valuing a condo unit is a three dimensional problem; i.e., *the height of the unit* and the *height of the building* are important price determining characteristics. In general, constructing constant quality condominium price indexes is much more difficult than constructing house price indexes.**

The Data (1)

- **Our basic data set is on sales of condominium units located in the central area of Tokyo over the 61 quarters starting at the first quarter of 2000 and ending at the first quarter of 2015.**

- **There were a total of 3,232 observations (after range deletions) in our sample of sales of condo units in Tokyo.**

Tokyo Special District:
Area: 626.70 km^2
Population: 9,256,625

The Data (2)

- **V = The value of the sale of the condo unit in 10,000 Yen;**
- **S = Structure area (floor space area) for the *unit*;**
- **TS = Floor space area for the *entire building*;**
- **TL = Lot area for the entire structure *in units* of meters squared;**
- **A = Age of the structure in years;**
- **H = The story of the unit; i.e., the height of the unit that was sold;**
- **TH = The total number of stories in the building;**
- **N = The number of units in the entire building;**

- **NB = Number of bedrooms in the unit;**
- **TW = Walking time in minutes to the nearest subway station;**
- **TT = Subway running time in minutes to the Tokyo station from the nearest station during the day (not early morning or night);**
- **SCR=Reinforced concrete construction dummy variable;**
- **SOUTH=Dummy variable.**

The Data (3)

- **In addition to the above variables, we also have information on which Ward of Tokyo the sales took place. We used this information to create ward dummy variables, $D_{W,tn,j}$.**

- **Ward 1 = Sumida; Ward 2 = Koto; Ward 3 = Kita; Ward 4 = Arakawa; Ward 5 = Itabashi; Ward 6= Nerima; Ward 7 = Adachi; Ward 8 = Katsushika and Ward 9 = Edogawa.**

- **In order to reduce multicollinearity between the various independent variables listed above (and to achieve consistency with national accounts data), we assumed that the value of a new structure in any quarter is proportional to a Construction Cost Price Index for Tokyo.**

→We denote the value of this index during quarter t as p_{St}.

The Basic Builder's Model (1)

- **The *builder's model* for valuing a residential property postulates that the value of a residential property is the sum of two components: the value of the land which the structure sits on plus the value of the residential structure.**

- **The total cost of the property after the structure is completed will be equal to the floor space area of the structure, say S square meters, times the building cost per square meter, β say, plus the cost of the land, which will be equal to the cost per square meter, α say, times the area of the land site, L.**

(1) $V_{tn} = \alpha_t L_{tn} + \beta_t S_{tn} + \varepsilon_{tn}$; $t = 1,\ldots,61$; $n = 1,\ldots,N(t)$.

The Builder's Model (2)

- **For older structures, we modify eq (1) and allow for *geometric depreciation* of the structure:**

(2) $V_{tn} = \alpha_t L_{tn} + \beta_t (1 - \delta_t)^{A(t,n)} S_{tn} + \varepsilon_{tn}$;

where the parameter δ_t reflects the *net geometric depreciation rate* as the structure ages one additional period and

- **L_{tn} is the unit's share of the total land plot area of the structure (how do we determine this?), α_t is the price of land (per meter squared), β_t is the price of condo floor space (per meter squared), A(t,n) is the age of the structure in years and S_{tn} is the floor space of the unit (in square meters).**
- **δ_t is regarded as a *net depreciation rate* because it is equal to a "true" gross structure depreciation rate less an average renovations appreciation rate. (We do not have information on renovation expenditures).**

Problems with the Builder's Model

- **There are at least two major problems with the hedonic regression model defined by (2):**

(i) The multicollinearity problem and

(ii) The problem of imputing an appropriate share of the total land area to a particular condominium unit.

- **Experience has shown that it is usually not possible to estimate sensible land and structure prices in a hedonic regression like that defined by (2) due to the multicollinearity between lot size and structure size. Thus we assume that the price of new structures is proportional to an official index of condominium building costs, p_{St}.**
- **Thus we replace β_t in (2) by βp_{St} for t = 1,...,61. This reduces the number of free parameters in the model by 60.**

The Land Share Imputation Problem

- There are two simple methods for constructing an appropriate land share:

(i) Use the unit's share of floor space to total structure floor space or

(ii) Simply use 1/N as the share where N is the total number of units in the building.

- Thus define the following two methods for making land imputations for unit n in period t:

(3) $L_{Stn} \equiv (S_{tn}/TS_{tn})TL_{tn}$; $L_{Ntn} \equiv (1/N_{tn})TL_{tn}$; $t = 1,\ldots,61; n = 1,\ldots,N(t)$

- where S_{tn} is the floor space area of unit n in period t, TS_{tn} is the total building floor space area, TL_{tn} is the total land area of the building and N_{tn} is the total number of units in the building for unit n sold in period t. The first method of land share imputation is used by the Japanese land tax authorities. The second method of imputation implicitly assumes that each unit can enjoy the use of the entire land area and so an equal share of land for each unit seems "fair".

A Problem with the First Method of Imputation

- The shares S_{tn}/TS_{tn}, if available for every unit in the building, would add up to a number less than one because the unit floor space areas, S_{tn}, if summed over all units in the building, add up to *privately owned floor space* which is less than total building floor space TS_{tn}.
- **Total building floor space includes halls, elevators, storage space, furnace rooms and other "public" floor space.**

- An approximation to total building privately owned floor space for observation n in period t is $\underline{N_{tn}S_{tn}}$.
- Thus an imperfect estimate of the ratio of privately owned floor space to total floor space for unit n in period t is $N_{tn}S_{tn}/TS_{tn}$.

- **The sample wide average of these ratios was <u>0.899</u>. Thus to account for shared structure space, we replaced the owned floor space variable in equation (2), S_{tn}, by $\underline{(1/0.899)S_{tn} = (1.1)S_{tn}}$.**

Preliminary Regressions using the Two Methods for Making the Land Share Imputations

- **In order to get preliminary land price estimates, we substituted the land estimates defined by (3) into the regression model (2), we replaced the β_t by βp_{St}, the S_{tn} by (1) S_{tn} and we assumed that the annual geometric depreciation rate δ_t was equal to 0.03. The resulting linear regression models become the models defined by (4) and (5) below:**

$$(4)\ V_{tn} = \alpha_t L_{Stn} + (1.1)\beta p_{St}(1 - 0.03)^{A(t,n)} S_{tn} + \varepsilon_{tn} ;$$

$$(5)\ V_{tn} = \alpha_t L_{Ntn} + (1.1)\beta p_{St}(1 - 0.03)^{A(t,n)} S_{tn} + \varepsilon_{tn} .$$

- **Thus we have 3,232 degrees of freedom to estimate 61 land price parameters α_t and one structure quality parameter β for a total of 62 parameters for each of the models defined by (4) and (5).**
- **The R^2 for the models defined by (4) and (5) were only 0.5894 and 0.5863.**

A Problem with Our *Preliminary Regressions*

- **The estimates for β were 2.164 and 2.154 respectively which was totally *unsatisfactory* because these parameters should have been close to unity.**

- **Moreover the land price indexes that these regression models generated were subject to excessive volatility (due to the very high estimates for the structure quality parameter, β).**

- **In order to deal with the problem of too high estimates of β, we decided not to estimate it.**

- **Moreover, we temporarily put aside the problem of jointly determining land and structure value to concentrate on determining sensible constant quality land prices. Once sensible land prices have been determined, we will then return to the problem of simultaneously determining land and structure values and constant quality price indexes.**

Imputed Land Value becomes our Dependent Variable

- **In sections 4-10, we assumed that the structure value for unit n in period t, V_{Stn}, is defined as follows:**

$$(6)\ V_{Stn} \equiv (1.1)p_{St}(1 - 0.03)^{A(t,n)}S_{tn}; \qquad t = 1,\ldots,61;\ n = 1,\ldots,N(t).$$

- **Once the imputed value of the structure has been defined by (6), we define the imputed land value for condo n in period t, V_{Ltn}, by subtracting the imputed structure value from the total value of the condo unit, which is V_{tn}:**

$$(7)\ V_{Ltn} \equiv V_{tn} - V_{Stn}; \qquad t = 1,\ldots,61;\ n = 1,\ldots,N(t).$$

- **Thus in the following 7 sections, we use V_{Ltn} as our dependent variable and we will attempt to explain variations in these imputed land values in terms of the property characteristics.**
- **However, in the end, we will return to using property value as the dependent variable and we will estimate the depreciation rate.**

A Preliminary Land Value Regression

- For now, we will use the first land measure in (3) as our estimate of the share of total land that is imputed to unit n sold in period t; i.e., unit n's share of land in period t is measured as $L_{Stn} = (S_{tn}/TS_{tn})TL_{tn}$.
- We will estimate the following preliminary linear regression model where imputed land value V_{Ltn} has replaced total value V_{tn} as the dependent variable:

$$(8)\ V_{Ltn} = \alpha_t L_{Stn} + \varepsilon_{tn}\ ; \qquad t = 1,\ldots,61;\ n = 1,\ldots,N(t).$$

- The above simple linear regression model has 61 land price parameters α_t to be estimated.
- The R^2 between the observed and predicted variables was only 0.0064 and the log likelihood was −25913.6. These results are hardly stellar but on a positive note, the resulting land price index was reasonably behaved.

The Introduction of *Ward Dummy* Variables

- In order to take into account possible neighbourhood effects on the price of land, we introduce ***ward dummy variables***, $D_{W,tn,j}$, into the hedonic regression (8). These 9 dummy variables are defined as follows:

(9) $D_{W,tn,j} \equiv 1$ if observation n in period t is in Ward j of Tokyo;

$\equiv 0$ if observation n in period t is *not* in Ward j of Tokyo.

- We now modify the model defined by (8) to allow **the *level* of land prices to differ across the 9 Wards**. The new nonlinear regression model is the following one:

(10) $V_{Ltn} = \alpha_t(\sum_{j=1}^{9} \omega_j D_{W,tn,j})L_{Stn} + \varepsilon_{tn}$.

- We need to impose at least one **identifying normalization** on the above parameters:

(11) $\alpha_1 \equiv 1$.

- The R^2 for this model turned out to be 0.1237 and the log likelihood (LL) was –25433.0, a **big increase of 480.6** over the preliminary linear regression (8).

Building Height as an Explanatory Variable (1)

- **It is likely that the height of the building increases the value of the land plot supporting the building, all else equal.**
- **In our sample of condo sales, the height of the building (the TH variable) ranged from 3 stories to 22 stories. However, there were very few observations for the last 7 height categories. Thus we collapsed the last seven height categories into a single category 14 and the remaining 13 height categories corresponded to building heights of 3 to 15 stories. Thus we define the building height dummy variables, $D_{TH,tn,h}$, as follows:**

(12) $D_{TH,tn,h}$ ≡ 1 if observation n in period t is in building height category h;

≡ 0 if observation n in period t is *not* in building height category h.

- **The new nonlinear regression model is the following one:**

Building Height as an Explanatory Variable (2)

(13) $V_{Ltn} = \alpha_t(\sum_{j=1}^{9} \omega_j D_{W,tn,j})(\sum_{h=1}^{14} \chi_h D_{TH,tn,h})L_{Stn} + \varepsilon_{tn}$

- Comparing the models defined by equations (10) and (13), it can be seen that we have added an additional 14 *building height parameters*, $\chi_1,\ldots,\chi_{14}$, to the model defined by (10).
- However, looking at (13), it can be seen that the 61 land price parameters (the α_t), the 9 ward parameters (the ω_j) and the 14 building height parameters (the χ_h) cannot all be identified. Thus we imposed the following identifying normalizations on these parameters:

(14) $\alpha_1 \equiv 1$; $\chi_1 \equiv 1$.

- The R^2 for this model turned out to be 0.2849 and the log likelihood was –24831.8, a big increase of 601.2 over the LL of the model defined by (10) for the addition of 13 new parameters.
- Thus the height of the building is a very significant determinant of Tokyo condominium land prices.

The Height of the Unit as an Explanatory Variable

- **The higher up a unit is, the better is the view on average and so we would expect the price of the unit would increase all else equal.**
- **The quality of the structure probably does not increase as the height of the unit increases so it seems reasonable to impute the height premium as an adjustment to the land price component of the unit.**
- **Thus the new nonlinear regression model is the following one (the previous normalizations (15) were also imposed):**

$$(15)\ V_{Ltn} = \alpha_t(\textstyle\sum_{j=1}^{9} \omega_j D_{W,tn,j})(\sum_{h=1}^{14} \chi_h D_{TH,tn,h})(1+\gamma(H_{tn}-3))L_{Stn} + \varepsilon_{tn}\ .$$

The estimated value for γ turned out to be $\gamma^* = 0.0225$ ($t = 6.44$). Thus the imputed land value of a unit increases by 2.25% for each story above the threshold level of 3. (LL increase was 26)

A More General Method of Land Imputation

- **We set the land imputation for unit n in period t, L_{tn}, equal to a *weighted average* of the two imputation methods and estimate the best fitting weight, λ. Thus we define:**

(16) $L_{tn}(\lambda) = [\lambda(S_{tn}/TS_{tn}) + (1-\lambda)(1/N_{tn})]TL_{tn}$

- **The new nonlinear regression model is the following one:**

(17) $V_{Ltn} = \alpha_t(\sum_{j=1}^{9} \omega_j D_{W,tn,j})(\sum_{h=1}^{14} \chi_h D_{TH,tn,h})(1+\gamma(H_{tn}-3))L_{tn}(\lambda) + \varepsilon_{tn}$;

- **The R^2 was 0.3021 and the LL was –24644.8, a big increase of 161.0 over the previous model for the addition of one new parameter.**
- **The estimated λ was $\lambda^* = 0.3636$ (t = 9.84) which is the weight for the floor space allocation method and the weight for the number of units in the building was 0.6364.**

The Number of Units in the Building as an Explanatory Variable

- **Conditional on the land area of the building, we expect the sold unit's land imputation value to increase as the number of units in the building increases.**
- **The range of the number of units in the building, N_{tn}, in our sample was from 11 to 154 units.**
- **Thus we introduce the term $1+\kappa(N_{tn}-11)$ as an explanatory term in the nonlinear regression. The new parameter κ is the percentage increase in the unit's imputed value of land as the number of units in the building grows by one unit.**
- **The new nonlinear regression model is the following one:**

$$(18)\ V_{Ltn} = \alpha_t(\textstyle\sum_{j=1}^{9} \omega_j D_{W,tn,j})(\sum_{h=1}^{14} \chi_h D_{TH,tn,h})(1+\gamma(H_{tn}-3)) * (1+\kappa(N_{tn}-11))L_{tn}(\lambda) + \varepsilon_{tn}\ .$$

- **The R^2 for this model was 0.3081 and the LL was –24604.4, a substantial increase of 40.4 over the previous model.**

Subway Travel Times and *Facing South* as Explanatory Variables (1)

- **There are three additional explanatory variables in our data set that may affect the price of land.**
- **Recall that TW was defined as walking time in minutes to the nearest subway station; TT as the subway running time in minutes to the Tokyo station from the nearest station and the SOUTH dummy variable is equal to 1 if the unit faces south and 0 otherwise.**
- **Let $D_{S,tn,2}$ equal the SOUTH dummy variable for sale n in quarter t. Define $D_{S,tn,2} = 1 - D_{S,tn,1}$.**
- **TW ranges from 1 to 19 minutes while TT ranges from 12 to 48 minutes.**
- **These new variables are inserted into the nonlinear regression model (21) in the following manner:**

Subway Travel Times and *Facing South* as Explanatory Variables (2)

(23) $V_{Ltn} = \alpha_t(\sum_{j=1}^{9} \omega_j D_{W,tn,j})(\sum_{h=1}^{14} \chi_h D_{TH,tn,h})(\sum_{m=1}^{10} \mu_m D_{EL,tn,m})$

$\times(\phi_1 D_{S,tn,1} + \phi_2 D_{S,tn,2})(1+\gamma(H_{tn}-3))(1+\kappa(N_{tn}-11))$

$\times(1+\eta(TW_{tn}-1))(1+\theta(TT_{tn}-12))L_{tn}(\lambda) + \varepsilon_{tn}$;

(24) $\alpha_1 \equiv 1; \chi_1 \equiv 1; \mu_1 \equiv 1; \phi_1 \equiv 1.$

- The R^2 for this model turned out to be 0.6308 and the log likelihood was –23178.30, a huge increase of 405.8 over the LL of the previous model for the addition of 3 new parameters.
- The estimated facing south parameter is $\phi_2^* = 1.0294$ (t = 120.6) so the land value of a condo unit that faces south increases by 2.94%.
- The walking to the subway parameter turns out to be $\eta^* = -0.0176$ (t = –26.7) so that an extra minute of walking time reduces the land value component of the condo by 1.76%. The travel time to the Tokyo Central Station parameter is $\theta^* = -0.0128$ (t = –27.4) so that an extra minute of travel time reduces the land value component of the condo by 1.28%. These are reasonable numbers.

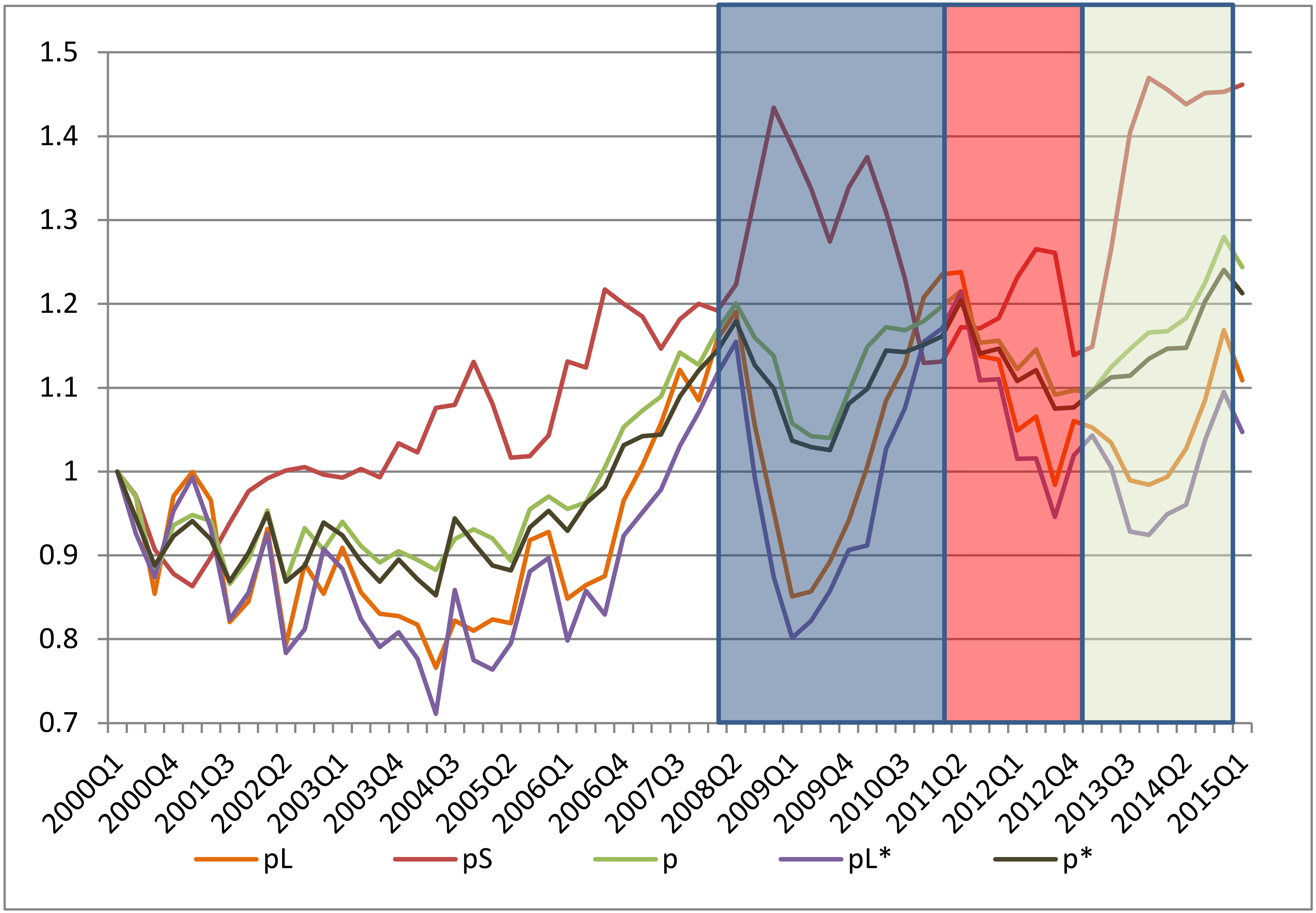
Using the *Selling Price* as the Dependent Variable

- **We switch from imputed land value V_{Ltn} as the dependent variable in the regressions to the selling price of the property, V_{tn}.**
- **We introduced the number of *bedrooms variable*, NB_{tn}, and the *reinforced concrete construction* SCR_{nt} dummy variable as quality adjusters for the value of the structure. The details are omitted.**

$$(26)\quad V_{tn} = \alpha_t(\Sigma_{j=1}^{9}\,\omega_j D_{W,tn,j})(\Sigma_{h=1}^{14}\,\chi_h D_{TH,tn,h})(\Sigma_{m=1}^{10}\,\mu_m D_{EL,tn,m}) \times(\phi_1 D_{S,tn,1}+\phi_2 D_{S,tn,2})(1+\gamma(H_{tn}-3))(1+\kappa(N_{tn}-11)) \times(1+\eta(TW_{tn}-1))(1+\theta(TT_{tn}-12))L_{tn}(\lambda) + (1.1)p_{St}(1-\delta)^{A(t,n)}(1+\sigma SRC_{tn})(\Sigma_{i=1}^{3}\,\rho_i D_{B,tn,i})S_{tn} + \varepsilon_{tn};$$

- **Basically, we now estimate the depreciation rate instead of assuming that it equals 3%.**
- **The R^2 for this new model turned out to be 0.8190 and the log likelihood was –23164.33. (Not comparable with (21) LL).**
- **The estimated depreciation rate was $\delta^* = 0.0367$ (t = 27.1). This estimated annual depreciation rate of 3.67% is higher than our earlier assumed rate of 3.00%.**

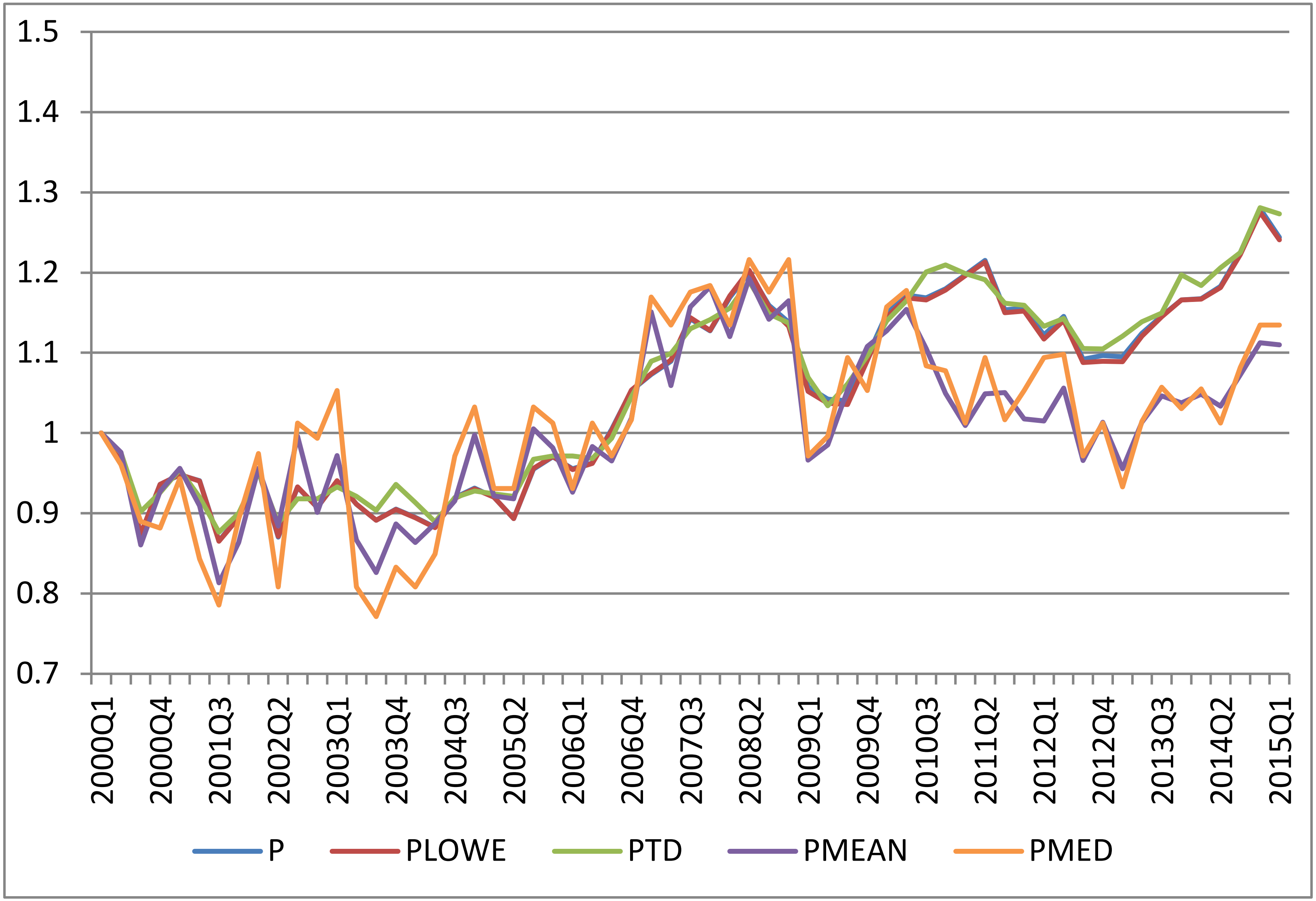
Land, Structure and Property Price Indexes p_{Lt}, p_{St} and p_t for the Section 12 Model and Land and Property Price Indexes p_{Lt}^* and p_t^* for the Section 13 Model



Comparison of the Section 12 Price Index with other Condo Price Indexes (1)

- **The price indexes for land, structures and the entire property on the previous slide were for sales of condo units. But for national accounts purposes, we need in particular, an index for the stock of land used to support condo units.**
- **We can form an approximation to the stock of condo units in our 9 wards by summing over all the units sold during the 61 periods in our sample; i.e., we replace *sales weights* by approximate *stock weights*. The resulting land and overall price indexes are *Lowe indexes* and they were very close to our Sales price index counterparts.**
- **The Overall Section 12 Sales Price Index p_t, the Lowe Index p_{LOWEt}, a Traditional Time Dummy Hedonic Regression Sales Price Index p_{TDt} and the Quarterly Mean and Median Price Indexes of Sales, p_{MEANt} and p_{MEDt}, are shown in the following figure.**

Comparison of the Section 12 Price Index with other Condo Price Indexes (2)



Conclusion

- **Our nonlinear regression approach led to an estimated geometric <u>depreciation rate for Tokyo apartment buildings of about 3.6% per year</u>, which seems reasonable. (<u>1.68%</u> from Traditional Time dummy hedonic.)**
- **Our preferred overall price index for condo sales was virtually identical to the corresponding Lowe index which provides an approximation to a price index for the stock of condo units in Tokyo.**
- **<u>Means and median indexes of condo sales tend to have a downward bias due to their neglect of net depreciation of the structure.</u>**
- **Traditional time dummy hedonic regressions can generate reasonable overall price indexes for condo sales. However, if the estimated age coefficient is large and positive, <u>the resulting time dummy price index is likely to have a substantial downward bias</u>.**
- **Our method does lead to a reasonable decomposition of condo property prices into land and structure components.**

Our Future Works

- **New estimation method for Transaction Based *<u>Commercial Property Price Indexes</u>*.**

eurostat

HANDBOOK ON COMMERCIAL PROPERTY PRICE INDICATORS