Hitotsubashi-RIETI International Workshop on Real Estate and the Macro Economy

Alternative Land Price Indexes for Commercial Properties in Tokyo

Research Institute of Economy, Trade and Industry, Tokyo

December 14, 2017

Chihiro Shimizu

(Nihon University & MIT)

with

W. Erwin Diewert

(The University of British Columbia)

LEVERAGING SCIENCE, DEVELOPING INNOVATION

Introduction: Commercial Property and SNA

- We will review some of the problems associated with the construction of price indexes for commercial properties.
- Property price indexes are required for the stocks of commercial properties in the Balance Sheet Accounts of the country.
- Related service price indexes for the land and structure input components of a commercial property are required in the Production Accounts of the country if the Multifactor Productivity of the Commercial Property Industry is calculated.
- We will mainly focus on existing methods for constructing an overall Commercial Property Price Index (CPPI).
- Many methods are biased (due to their neglect of *depreciation*) but more importantly, most methods are not able to provide separate land and structure subindexes.
- A class of hedonic regression models that deals adequately with these problems will be discussed in some detail.

The Measurement of Commercial Property Outputs

- **Typically, commercial buildings in a particular location are classified into four broad groups:**
- **Offices;**
- **Retail sales;**
- **Industrial (factories, warehouses, repair facilities, etc.);**
- **Residential apartments.**
- **Buildings could be further subdivided according to the type of construction, the location of the building and other characteristics.**
- **The outputs produced by an office or retail building consist primarily of the rental or leasing of individual units of floor space.**
- **The total floor space rented will generally be well below the total floor space of the building since some space will be taken up by hallways, utility rooms, caretaker and managerial offices.**
- **When measuring outputs, rented space is what counts but later when we value the primary input services provided by the basic building structure, it is total floor space that matters.**

The Measurement of Outputs (continued)

- **In addition to leased or rented space, the building may make additional revenues from renting parking spaces and other miscellaneous sources of revenues.**
- **If a building is rented or leased to multiple business entities, it is preferable to collect statistics on the outputs produced by the building and the inputs used from the owners of the building rather than (partially) survey the occupants of the building. This will ensure that the productivity of the building will be properly measured.**
- **The basic output price concept is the rent per meter squared of space. Thus if the owners of a building collect rent from N tenants and the revenue collected in period t from tenant n is v_n^t and the floor space area occupied by tenant n is q_n^t, then the corresponding price is $p_n^t \equiv v_n^t/q_n^t$.**

The Measurement of Outputs (continued)

- **Thus price and quantity data for the outputs produced by a commercial building can be collected from the owners or managers of buildings and thus weighted output indexes of the usual type (Paasche, Laspeyres, Fisher, Törnqvist) can be calculated.**
- **In general, it is not wise to collect rents from a panel of tenants. Many rents and leases change the rental price when a new occupant is signed up. The new occupant will usually not know what rent the previous occupant paid and thus a rent index based on surveys of occupants will tend to show little change.**
- **Vacancies lead to a zero price and quantity problem; i.e., if part of a building is temporarily vacant, then the corresponding rent and output quantity becomes 0.**

Problems with the Measurement of Outputs

- **If we set the price for a vacant component of a building equal to 0 and apply the usual index number formulae assuming 0 prices and quantities for this component, the resulting index will tend to be biased. In fact, a Jevons, weighted Jevons or Törnqvist price index will not be well defined under these circumstances. How should we deal with the zero or missing price and quantity problem?**
- **As long as prices are all positive for the two periods being compared, the usual indexes can be evaluated. Thus one solution is to impute a positive price for the corresponding 0 quantities. Carrying forward the last positive price is an obvious solution.**
- **However, the carry forward solution is not satisfactory; it will tend to lead to an understatement of inflation and an overstatement of deflation.**
- **A better solution is to calculate a maximum overlap index.**

Problems with the Measurement of Outputs (cont)

- **Another problem that arises in the measurement of commercial property outputs is the problem of quality change.**
- **In particular, the aging of a structure can lead to a loss of utility for the occupants of the structure. This is almost certainly the case for rental apartments; tenants prefer new buildings to older ones in general. Thus in this case, the price series should be adjusted upward to account for this loss of utility. A simple solution is to take the period t price level, say it is equal to P^t, and divide it by $(1-\delta)$ where δ is an estimated (net) geometric depreciation rate. (The corresponding quantity is multiplied by $(1-\delta)$. The BLS performs this operation to its rent index.**
- **However, for some commercial uses, space is space and no depreciation adjustment is required. In this case, we have one hoss shay depreciation where the flow of services yielded by the structure is constant through the life of the structure.**

Problems with the Measurement of Outputs (cont)

- **Another problem that needs some discussion is the problem of possible quality improvements in new buildings as compared to previous structures.**
- **Thus historically, buildings have become "better" due to improvements in insulation, in wiring, the introduction of double and triple glazed windows and lighting. Just following existing buildings will not capture technological improvements in structures.**
- **Possible new improvements are associated with "greening" buildings; i.e., solar panels and heat pumps can be installed on new buildings (and older buildings can be retrofitted).**
- **Newer buildings may also be more earthquake and hurricane resistant.**
- **These types of technological improvements need to be taken into account. Hedonic regression techniques or engineering studies can be used to make these adjustments.**

Problems with the Measurement of Outputs (concluded)

- **The final output measurement problem we will discuss is the problem of what to do when a commercial property is not rented or leased to a third party but instead is used as part of the overall production of a firm; i.e., we have an Owner Occupied Commercial Property (OOCP).**
- **In this case, there are no external rental or lease prices so how can we measure the output of these internally used commercial properties?**
- **One way of proceeding is to construct user costs for the structure and the land that the structure sits on and use these user costs along with other variable costs to form an input aggregate. The output price of the property would be set equal to this aggregate input price.**
- **The value of the service flow output of the property would be set equal to the total property cost (which would include the user costs of land and the structure).**

The Construction of Variable Input Price Indexes

- **Examples of intermediate and labour nondurable inputs that are used to provide commercial building services include the following:**
- **Inputs used to heat the building such as fuel oil, coal and natural gas;**
- **Electricity inputs;**
- **Telecommunication inputs;**
- **Cleaning supplies;**
- **Janitorial, maintenance and repair inputs;**
- **Insurance services;**
- **Security and caretaker services and**
- **Managerial and legal services inputs.**

The Construction of Variable Input Price Indexes (concl.)

- There are 4 periodic input costs for which there are values but no obvious breakdown into price and quantity components. These four classes of ***value only nondurable variable input costs*** are as follows:
- Property tax payments;
- Business income tax payments;
- Property insurance payments and
- Direct and indirect charges for undertaking monetary transactions and holding bank balances.

Expenditures on these cost categories can be obtained but the decomposition of these nominal costs into price and quantity components is a tricky business which we will not go into in this talk.

- We now turn to the determination of capital costs.

The Decomposition of Property Asset Values into Land and Structure Components

- **Capital theory tells us that the asset value of a property is equal to the discounted cash flow that it is expected to generate.**
- **A property consists of a quantity of *land* bundled together with a *structure* that sits on the land. Once the structure is built, we have two fixed costs: one for the land and one for the structure. The cash flows are generated by both fixed assets so it is difficult to assign definite fractions of cash flow to the two assets.**
- **However, the structure depreciates whereas land does not. This will help us to identify separate asset values for the structure and land components of property value.**
- **In the remainder of this talk, we will show how three alternative sources of data can be used to accomplish a land and structure split for the asset value of a property.**
- **Once asset values for the land and structure components of a property have been determined, then user costs for these capital stock components can be calculated.**

Alternative Data Sources

- **There are at least three alternative data sources suggested in the literature that enable one to construct land and structure price indexes for commercial properties:**

  (i) sales transactions data;

  (ii) appraisal data for Real Estate Investment Trusts (REITs);

  (iii) assessed values of land for property taxation purposes.
- **We will utilize these three sources of data for commercial properties in Tokyo over 44 quarters covering the period Q1:2005 to Q4:2015 and compare the resulting land prices.**
- **We will also indicate how (net) depreciation rates for the structure component of a commercial property can be estimated using hedonic regression models.**
- **We will find problems with all three sources of data but in the end, we will favour the use of sales transactions data.**

Data Description

Diewert and Shimizu (2017) compiled the following three types of micro-data relating to commercial properties in the Tokyo office market:

- **(i) the transaction price data compiled by the Japanese Ministry of Land, Infrastructure, Transport and Tourism;**
- **(ii) the appraisal prices periodically determined in the Tokyo office REIT market; and**
- **(iii) the "official land prices" surveyed by the Japanese Ministry of Land, Infrastructure, Transport and Tourism since 1970.**

Official land prices are based on appraisals that are released on January 1st of each year. Thus the resulting index will be an annual one. As official land prices are exclusively based on surveys of land prices, they do not include structure prices.

Data Description (continued)

The Table below lists our variables from the 3 sources.

	MLIT	REIT	OLP
V : Selling Price of Office Building (million yen)	394.18 (337.76)	6686.60 (4055.60)	1264.3 (1304.1)
S : Structure Floor Area (m^2)	834.00 (535.19)	8509.70 (5463.90)	-
L : Land Area (m^2)	239.27 (135.08)	1802.10 (1580.20)	229.94 (217.18)
H : Total Number of Stories	5.75 (2.14)	10.12 (3.30)	-
A : Age (years)	24.23 (10.61)	19.14 (6.80)	-
DS : Distance to Nearest Station (meters)	387.65 (238.45)	308.29 (170.04)	347.24 (254.79)
TT : Time to Tokyo Station (minutes)	19.63 (8.23)	15.88 (5.10)	21.74 (8.54)
PS : Structure Construction Price per m^2 (million yen)	0.2347 (0.0103)	0.2359 (0.0102)	-
Number of Observations	1,907	1,804	6,242

( ): Standard deviation

The Builder's Model Using the MLIT Transactions Data

- **The builder's model for valuing a commercial property postulates that the value of a commercial property is the sum of two components: the value of the land which the structure sits on plus the value of the commercial structure.**
- **In order to justify the model, consider a property developer who builds a structure on a particular property.**
- **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_t during quarter or year t, plus the cost of the land, which will be equal to the cost per square meter, α_t during quarter or year t, times the area of the land site, L.**
- **Now think of a sample of properties of the same general type, which have prices or values V_{tn} in period t and structure areas S_{tn} and land areas L_{tn} for $n = 1,\ldots,N(t)$ where N(t) is the number of observations in period t.**

The Builder's Model Using MLIT Data (cont.)

- **Assume that these prices are equal to the sum of the land and structure costs plus error terms ε_{tn} which we assume are independently normally distributed with zero means and constant variances. This leads to the following hedonic regression model for period t where the α_t and β_t are the parameters to be estimated in the regression:**

(1) $V_{tn} = \alpha_t L_{tn} + \beta_t S_{tn} + \varepsilon_{tn}$; $\quad t = 1,\ldots,44;\ n = 1,\ldots,N(t).$

- **Note that the two characteristics in our simple model are the quantities of land L_{tn} and the quantities of structure floor space S_{tn} associated with property n in period t and the two constant quality prices in period t are the price of a square meter of land α_t and the price of a square meter of structure floor space β_t.**

The Builder's Model Using MLIT Data (cont.)

- **The hedonic regression model defined by (1) applies to new structures. But it is likely that a model that is similar to (1) applies to older structures as well. Older structures will be worth less than newer structures due to the depreciation of the structure.**
- **Assuming that we have information on the age of the structure n at time t, say A(t,n), and assuming a geometric (or declining balance) depreciation model, a more realistic model is the following basic builder's model:**

$$(2)\ V_{tn} = \alpha_t L_{tn} + \beta_t (1 - \delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}\ ; \quad t = 1,\ldots,44;\ n = 1,\ldots,N(t)$$

- **where the parameter δ reflects the net geometric depreciation rate as the structure ages one additional period.**
- **Thus if the age of the structure is measured in years, we would expect an annual *net* depreciation rate to be between 2 to 3%.**

The Builder's Model Using MLIT Data (cont.)

- There is a major problem with the hedonic regression model defined by (2): The multicollinearity problem.
- 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 assumed that the price of new structures is equal to an official measure of commercial building costs (per square meter of building structure), p_{St}. Thus we replaced β_t in (2) by p_{St} for t = 1,...,44. This reduces the number of free parameters in the model by 44.
- Experience has also shown that it is difficult to estimate the depreciation rate before obtaining quality adjusted land prices.
- Thus in order to get preliminary land price estimates, we temporarily assumed that the annual geometric depreciation rate δ in equation 2 was equal to 0.025.

The Builder's Model Using MLIT Data: Model 1

- **The resulting regression model becomes the model defined by (3) below:**

(3) $V_{tn} = \alpha_t L_{tn} + p_{St}(1 - 0.025)^{A(t,n)}S_{tn} + \varepsilon_{tn}$; $t=1,\ldots,44$; $n=1,\ldots,N(t)$.

- **The final log likelihood for this Model 1 was –13328.15 and the R^2 was 0.4003.**
- **In order to take into account possible neighbourhood effects on the price of land, we introduce ward dummy variables, $D_{W,tnj}$, into the hedonic regression (3). There are 23 wards in Tokyo special district.**
- **We made 23 ward or locational dummy variables. These 23 dummy variables are defined as follows:**

(4) $D_{W,tnj} \equiv 1$ 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.

The Builder's Model Using MLIT Data: Preliminary Model 2

- **We now modify the model defined by (3) to allow the *level* of land prices to differ across the Wards. The new nonlinear regression model is the following one:**

$$(5)\ V_{tn} = \alpha_t(\textstyle\sum_{j=1}^{23} \omega_j D_{W,tnj})L_{tn} + p_{St}(1 - 0.025)^{A(t,n)}S_{tn} + \varepsilon_{tn}\ ; \quad t = 1,\ldots,44;\ n = 1,\ldots,N(t).$$

- **For identification of the parameters, we impose the following normalization on our coefficients:**

$$(6)\ \alpha_1 = 1.$$

- **The final log likelihood for the model defined by (5) and (6) was –12956.60 (increase of 371.55) and the R^2 was 0.5925.**
- **However, many of the wards had only a small number of observations and thus it is unlikely that our estimated ω_j for these wards would be very accurate.**
- **In order to deal with the problem of too few observations in many wards, we used the results of the above model to group the 23 wards into 4 Combined Wards based on their estimated ω_j coefficients.**

The Builder's Model Using MLIT Data: Model 2

- We reran the nonlinear regression model defined by (5) and (6) using just the 4 Combined Wards (call this Model 2) and the resulting log likelihood was –12974.31 and the R^2 was 0.5850.
- Thus combining the original wards into grouped wards resulted in a small loss of fit and a decrease in log likelihood of 17.71 when we decreased the number of ward parameters by 19.
- We regarded this loss of fit as an acceptable tradeoff.
- In our next model, we introduce some nonlinearities into the pricing of the land area for each property.
- The land plot areas in our sample of properties ran from 100 to 790 meters squared.
- Up to this point, we have assumed that land plots in the same grouped ward sell at a constant price per m^2 of lot area.

The Builder's Model Using MLIT Data: Model 3

- **It is likely that very large lots sell at an average price that is below the average price of medium sized lots.**
- **We initially divided up our 1907 observations into 7 groups of observations based on their lot size. The Group 1 properties had lots less than 150 m^2, the Group 2 properties had lots greater than or equal to 150 m^2 and less than 200 m^2, the Group 3 properties had lots greater than or equal to 200 m^2 and less than 300 m^2, ... and the Group 7 properties had lots greater than or equal to 600 m^2. However, there were very few observations in Groups 4 to 7 so we added these groups to Group 4.**
- **For each observation n in period t, we define the 4 *land dummy variables*, $D_{L,tnk}$, for k = 1,...,4 as follows:**

(7) $D_{L,tnk} \equiv 1$ if observation tn has land area that belongs to group k;

$\equiv 0$ if observation tn has land area that does not belong to group k.

The Builder's Model Using MLIT Data: Model 3 (cont)

- **These dummy variables are used in the definition of the following piecewise linear function of L_{tn}, $f_L(L_{tn})$, defined as follows:**

$$(8)\ f_L(L_{tn}) \equiv D_{L,tn1}\lambda_1 L_{tn} + D_{L,tn2}[\lambda_1 L_1+\lambda_2(L_{tn}-L_1)]$$
$$+ D_{L,tn3}[\lambda_1 L_1+\lambda_2(L_2-L_1)+\lambda_3(L_{tn}-L_2)]$$
$$+ D_{L,tn4}[\lambda_1 L_1+\lambda_2(L_2-L_1)+\lambda_3(L_3-L_2) +\lambda_4(L_{tn}-L_3)]$$

- **where the λ_k are unknown parameters and $L_1 \equiv 150$, $L_2 \equiv 200$ and $L_3 \equiv 300$. The function $f_L(L_{tn})$ defines a relative valuation function for the land area of a commercial property as a function of the plot area.**
- **Basically, we are fitting a spline function on the land area.**
- **The new nonlinear regression model is the following one:**

$$(9)\ V_{tn} = \alpha_t(\textstyle\sum_{j=1}^{4} \omega_j D_{W,tnj})f_L(L_{tn}) + p_{St}(1-\delta)^{A(t,n)}S_{tn} + \varepsilon_{tn}\ ;$$
$$t = 1,\ldots,44;\ n = 1,\ldots,N(t).$$

The Builder's Model Using MLIT Data: Model 3 (concluded)

- We impose the following identification normalizations on the parameters for Model 3 defined by (9) and (10):

(10) $\alpha_1 \equiv 1; \lambda_1 \equiv 1.$

- Note that if we set all of the λ_k equal to unity, Model 3 collapses down to Model 2.
- The final log likelihood for Model 3 was an improvement of 59.65 over the final LL for Model 2 (for adding 3 new marginal price of land parameters) which is a highly significant increase.
- The R^2 increased to 0.6116 from the previous model R^2 of 0.5850.
- The parameter estimates turned out to be $\lambda_2 = 1.4297$, $\lambda_3 = 1.2772$ and $\lambda_4 = 0.2973$. These estimates indicate that the price of land increases initially as lot size increases but eventually decreases substantially as lot sized becomes large.

The Builder's Model Using MLIT Data: Model 4

- **The footprint of a building is the area of the land that directly supports the structure.**
- **An approximation to the footprint land for property n in period t is the total structure area S_{tn} divided by the total number of stories in the structure, H_{tn}.**
- **If we subtract footprint land from the total land area, TL_{tn}, we get excess land, EL_{tn} defined as follows:**

(11) $EL_{tn} \equiv L_{tn} - (S_{tn}/H_{tn})$; $t = 1,\ldots,44;\ n = 1,\ldots,N(t)$.

- **In our sample, excess land ranged from 1.083 m^2 to 562.58 m^2. We grouped our observations into 5 categories, depending on the amount of excess land that pertained to each observation. Group 1 consists of observations tn where1: $EL_{tn} < 50$; 2: observations such that $50 \le EL_{tn} < 100$; 3: $100 \le EL_{tn} < 150$; 4: $150 \le EL_{tn} < 300$; 5: $EL_{tn} \ge 300$. Now define the excess land dummy variables, $D_{EL,tnm}$, as follows: for $t = 1,\ldots,44$; $n = 1,\ldots,N(t)$; $m = 1,\ldots,5$:**

(12) $D_{EL,tnm} \equiv 1$ if observation n in period t is in excess land group m;

$\equiv 0$ if observation n in period t is not in excess land group m.

The Builder's Model Using MLIT Data: Model 4 (concluded)

- **As will be seen, in general, the more excess land a property possessed, the lower was the average per meter squared value of land for that property.**
- **The new Model 4 nonlinear regression model is:**

(13) $V_{tn} = \alpha_t(\sum_{j=1}^{4} \omega_j D_{W,tnj})(\sum_{m=1}^{5} \chi_m D_{EL,tnm}) f_L(L_{tn})$
$+ p_{St}(1 - \delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}$;

(14) $\alpha_1 \equiv 1; \lambda_1 \equiv 1; \chi_1 \equiv 1$ **(identifying normalizations).**

- **The final log likelihood for Model 4 was an improvement of 23.99 over the final LL for Model 3.**
- **The R^2 increased to 0.6207 from the previous model R^2 of 0.6116.**
- **The χ_m parameter estimates turned out to be $\chi_2 = 0.9173$, $\chi_3 = 0.7540$, $\chi_4 = 0.7234$ and $\chi_5 = 0.8611$.**
- **Thus excess land does reduce the average per meter price of land.**

The Builder's Model Using MLIT Data: Model 5

- 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 commercial property prices, the height of the building (the H variable) ranged from 3 stories to 14 stories.
- **Model 5** is the following nonlinear regression model (where H_{tn} is the number of stories of the structure for property n):

$$(17)\ V_{tn} = \alpha_t(\Sigma_{j=1}^{4}\omega_j D_{W,tnj})(\Sigma_{m=1}^{5}\chi_m D_{EL,tnm})(1+\mu(H_{tn}-3))f_L(L_{tn}) + p_{St}(1-\delta)^{A(t,n)}S_{tn} + \varepsilon_{tn}\ ;$$

- Not all of the parameters in (17) can be identified so we again impose the normalizations (14).
- The final log likelihood for Model 5 was –12685.19, a big improvement of **205.47** over the final log likelihood for Model 4 (for adding 1 new height parameters). The R^2 increased to **0.6923** from the Model 4 R^2 of 0.6207.

The Builder's Model Using MLIT Data: Model 6

- **The height parameter μ turned out to be 0.2358. Thus the land value of the property increased 23.58% for each extra story of structure. This is a very substantial height premium.**

- **Model 6 is the same as Model 5 except that we estimated the annual geometric depreciation rate δ instead of assuming that it was equal to 2.5%.**
- **The final log likelihood for Model 6 was –12680.66, an improvement of 4.53 over the final log likelihood for Model 5 (for adding 1 new parameter). (Not much improvement).**
- **The R^2 increased marginally to 0.6938 from the previous model R^2 of 0.6923.**
- **The estimated depreciation rate was 4.76% with a standard error of 0.009. This rate seems high!**

The Builder's Model Using MLIT Data: Model 7

- **Recall that we used building height as a quality adjustment factor for the land area of the property.**
- **In our next model, we will use building height as a quality adjustment factor for the structure component of the property.**
- **Recall that the 12 building height dummy variables $D_{H,tnh}$ were defined by (15) above for h = 3,4, ..., 14. Due to the small number of observations in the last 5 height categories, we combined these dummy variables into a single height category that included all buildings of height 10 to 14 stories; i.e., the new $D_{H,tn10}$ was defined as $\Sigma_{h=10}^{14} D_{H,tnh}$.**
- **Model 7 is defined as the following nonlinear regression model:**

$$(18)\ V_{tn} = \alpha_t(\Sigma_{j=1}^{4}\omega_j D_{W,tnj})(\Sigma_{m=1}^{5}\chi_m D_{EL,tnm})(1+\mu(H_{tn}-3))f_L(L_{tn}) + p_{St}(1-\delta)^{A(t,n)}(\Sigma_{h=3}^{10}\phi_h D_{H,tnh})S_{tn} + \varepsilon_{tn}\ ;$$

The Builder's Model Using MLIT Data: Model 7 (concluded)

- **In addition to the normalizations in (14), we also imposed the normalization $\phi_3 = 1$ in order to insure a reasonable split between structure and land values.**
- **The final log likelihood for Model 7 was – 12640.40, an improvement of 40.26 over the final log likelihood for Model 6 (for adding 7 new parameters).**
- **The R^2 increased to 0.7063 from the previous model R^2 of 0.6938.**
- **The estimated depreciation rate δ was 3.41% with a standard error of 0.0077. (This is a smaller std error than before).**
- **The estimated $\phi_4,...,\phi_{10}$ were equal to 1.11, 1.31, 1.32, 1.11, 1.83, 2.01 and 2.12 (recall that ϕ_3 was set equal to 1). Thus as the height of the structure increased, the quality adjusted quantity of the structure increased (except for buildings with 7 stories; i.e., ϕ_7 was less than ϕ_6).**

The Builder's Model with Multiple Geometric Depreciation Rates; Model 8

- **In the following model, we allowed the geometric depreciation rates to differ after each 10 year interval .**
- **For each observation n in period t, we define the 5 *age dummy variables*, $D_{A,tni}$, for i = 1,...,5 as follows:**

(19) $D_{A,tni} \equiv 1$ if observation tn has structure age that belongs to age group i; $\equiv 0$ if observation tn has structure age that does not belong to age group i.

- **These age dummy variables are used in the definition of the following *aging function*, $g_A(A_{tn})$, defined as follows:**

(20) $$g_A(A_{tn}) \equiv D_{A,tn1}(1-\delta_1)^{A(t,n)} + D_{A,tn2}(1-\delta_1)^{10}(1-\delta_2)^{(A(t,n)-10)}$$
$$+ D_{A,tn3}(1-\delta_1)^{10}(1-\delta_2)^{10}(1-\delta_3)^{(A(t,n)-20)}$$
$$+ D_{A,tn4}(1-\delta_1)^{10}(1-\delta_2)^{10}(1-\delta_3)^{10}(1-\delta_4)^{(A(t,n)-30)}$$
$$+ D_{A,tn5}(1-\delta_1)^{10}(1-\delta_2)^{10}(1-\delta_3)^{10}(1-\delta_4)^{10}(1-\delta_5)^{(A(t,n)-40)} .$$

The Builder's Model Using MLIT Data: Model 8 (concluded)

- **The new Model 8 nonlinear regression model is the following one:**

$$(21)V_{tn} = \alpha_t(\Sigma_{j=1}^{4}\omega_j D_{W,tnj})(\Sigma_{m=1}^{5}\chi_m D_{EL,tnm})(1+\mu(H_{tn}-3))f_L(L_{tn}) + p_{St}g_A(A_{tn})(\Sigma_{h=3}^{10}\phi_h D_{H,tnh})S_{tn} + \varepsilon_{tn}$$

- **We imposed the normalizations $\alpha_1 \equiv 1$, $\lambda_1 \equiv 1$, $\chi_1 \equiv 1$ and $\phi_3 \equiv 1$.**
- **Note that Model 8 collapses down to Model 7 if $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta$.**
- **The final log likelihood for Model 8 was –12631.21, an improvement of 9.19 over the final log likelihood for Model 7 (for adding 4 additional parameters).**
- **The R^2 increased to 0.7091 from the previous model R^2 of 0.7063.**
- **The estimated depreciation rates were as follows: $\delta_1 = 0.0487$ (0.0111), $\delta_2 = 0.0270$ (0.0097), $\delta_3 = 0.0096$ (0.0106), $\delta_4 = 0.0403$ (0.0154), $\delta_5 = -0.0319$ (0.0185), a negative rate for old buildings**

The Builder's Model Using MLIT Data: Model 9

- **DS is defined as the distance to the nearest subway station and TT as the subway running time in minutes to the Tokyo station from the nearest station.**
- **DS ranges from 0 to 1,500 meters while TT ranges from 1 to 48 minutes. Typically, as DS and TT increase, land value decreases.**
- **Model 9 introduces these new variables into the previous nonlinear regression model (21) in the following manner:**

$$(22)\, V_{tn} = \alpha_t(\Sigma_{j=1}{}^{4}\omega_j D_{W,tnj})(\Sigma_{m=1}{}^{5}\chi_m D_{EL,tnm})(1+\mu(H_{tn}-3))(1+\eta(DS_{tn}-0))(1+\theta(TT_{tn}-1))f_L(L_{tn}) + p_{St}g_A(A_{tn})(\Sigma_{h=3}{}^{10}\,\phi_h\, D_{H,tnh})S_{tn} + \varepsilon_{tn}\,;$$

- **Thus two new parameters, η and θ, are introduced.**
- **If these new parameters are both equal to 0, then Model 9 collapses down to Model 8.**

The Builder's Model Using MLIT Data: Model 9 (cont)

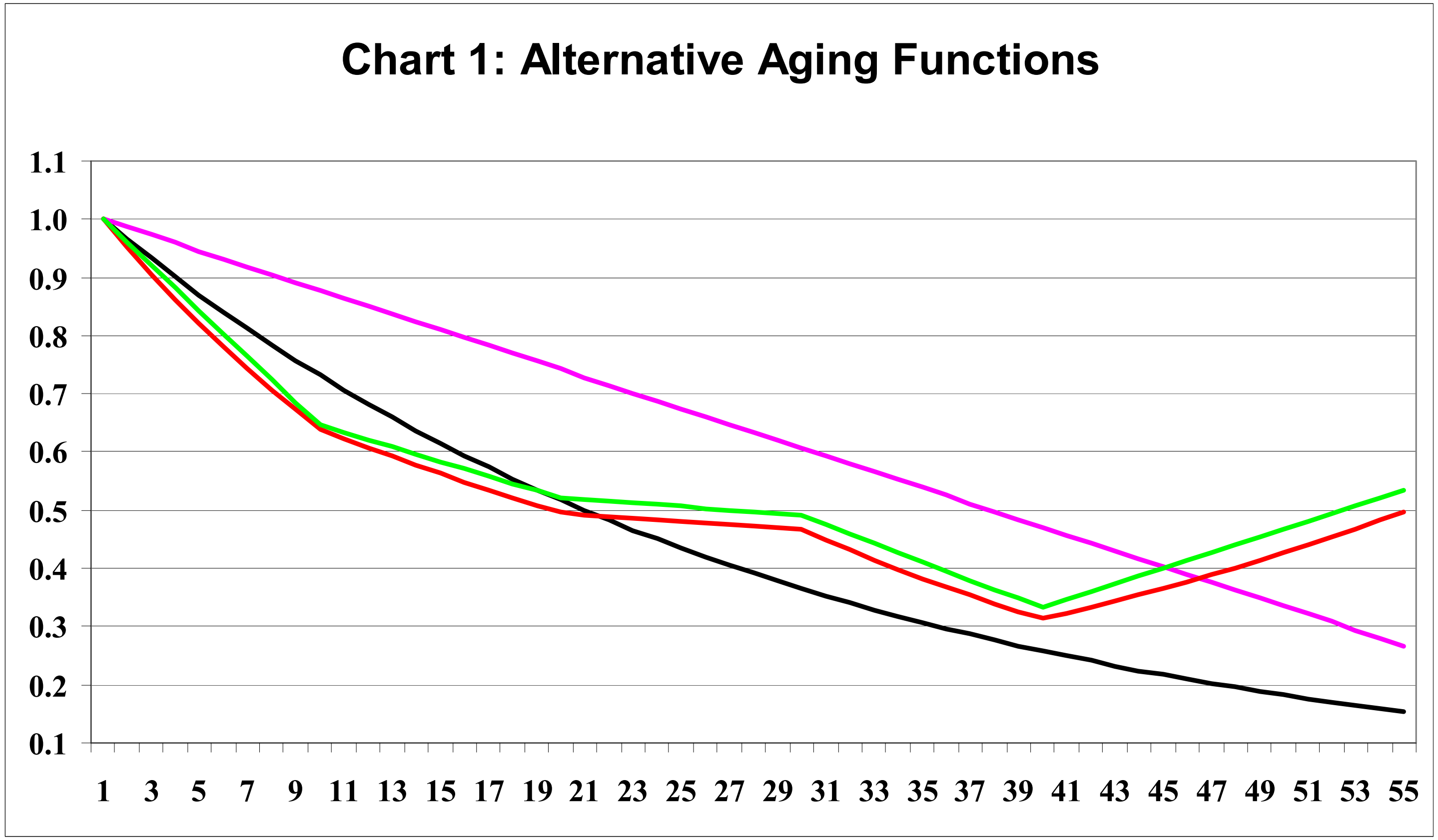
- **The final log likelihood for Model 9 was –12614.70, an improvement of 16.51 over the final log likelihood for Model 8 (for adding 2 additional parameters).**
- **The R^2 increased to 0.7142 from the previous model R^2 of 0.7091.**
- **The estimated walking distance parameter was $\eta = -0.00023$ (0.000066), which indicates that commercial property land value does tend to decrease as the walking distance to the nearest subway station increases.**
- **However, the estimated travel time to Tokyo Central Station parameter was $\theta = 0.0209$ (0.0053) which indicates that land value increases on average as the travel time to the central station increases, a relationship which was not anticipated.**
- **The estimated geometric depreciation rates were as follows: 4.84%, 2.52%, 0.60%, 3.89% and -3.12% for age 40+.**

The Straight Line and Piece-Wise Linear Depreciation Model

- **We also estimated a similar model with straight line depreciation.**
- **The estimated straight line depreciation rate was 1.36% per year. The R^2 for this Model 10 was 0.7078.**
- **We then estimated a piece-wise linear depreciation rate model with the same break points as our multiple geometric rates model. The R^2 for this Model 11 was 0.7143.**
- **The increase in Log Likelihood was 21.48 over Model 10.**
- **The estimated depreciation rates were as follows: 3.93%, 1.25%, 0.30%, 1.59% and -1.35% for age 40+.**
- **In the following slide, we show how structure value declines (at constant prices) due to the aging of the structure for the geometric and straight line models of depreciation and for their multiple rate generalizations.**

The Various Depreciation Models Compared

The top line is the straight line aging function. The green line is the multiple geometric rate function, the red line is the piece-wise linear depreciation aging function and the bottom black line is the single geometric rate aging function. The red and green lines are very close.

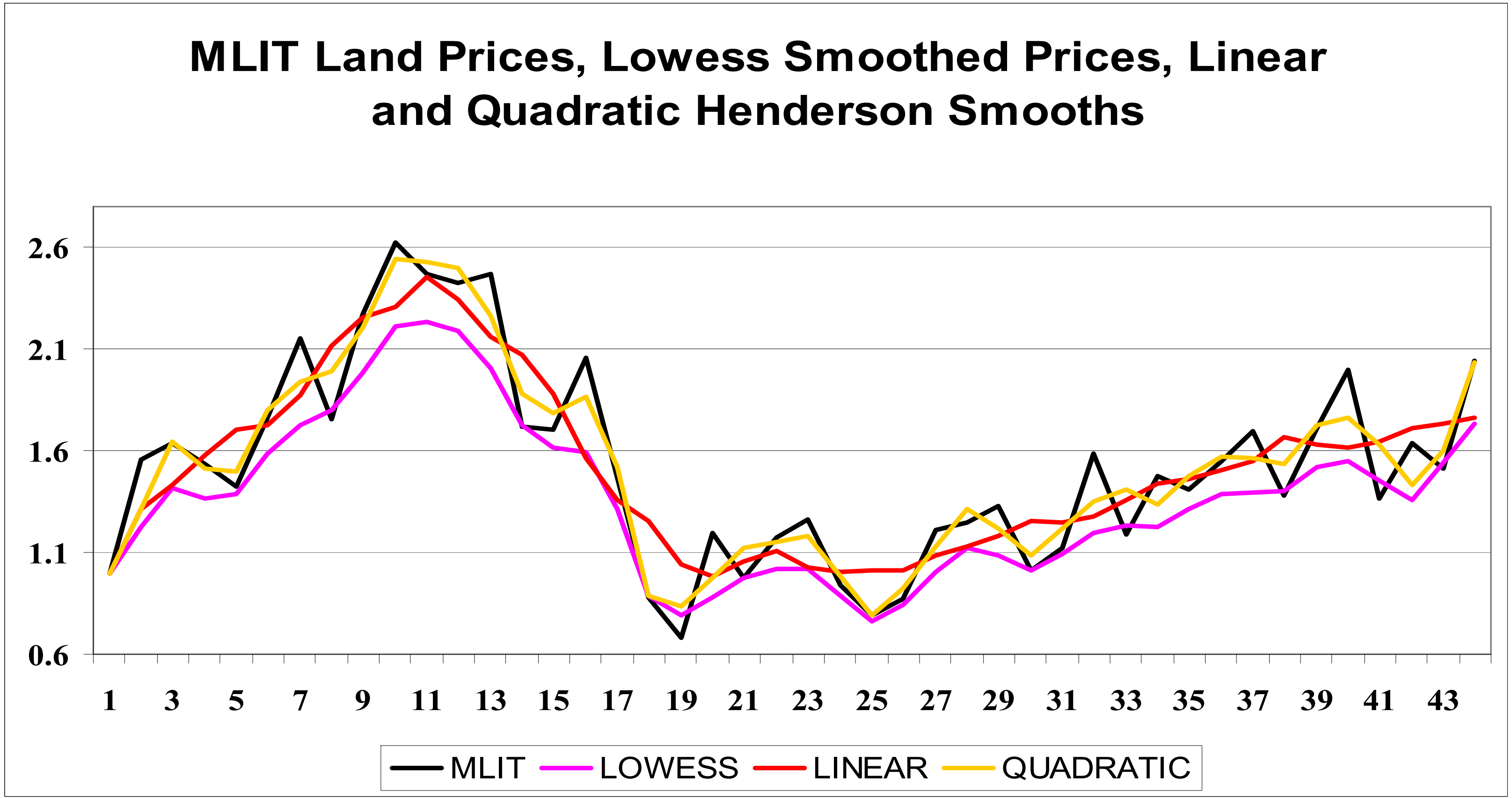


MLIT Land Prices and the Smoothing Problem

- **Once the hedonic regression model has been estimated, it is straightforward to compute the resulting land price series.**
- **However, due to the low number of transactions and the heterogeneity of the commercial office properties, the resulting index is not very smooth; see the next slide.**
- **Thus we followed the example of Ireland and looked at various smoothing methods to reduce the volatility of the index.**
- **The Lowess nonparametric smooth using Shazam is shown on the next slide. This approximation was not satisfactory; it was too low.**
- **Henderson (1916) was the first to realize that various moving average smoothers could be related to rolling window least squares regressions that would exactly reproduce a polynomial curve.**
- **Thus we applied his idea to derive the moving average weights that would be equivalent to fitting a linear (and also a quadratic) function to 5 consecutive quarters of a time series**

Smoothing the MLIT Hedonic Land Price Series

The Lowess nonparametric smoother is the purple line. The unsmoothed land price series is the black line. The quadratic smoother is the gold line (bit too wiggly) and the red line is our preferred linear smoother. The details are in the paper.



The Builder's Model Using Property Appraisal Data

- **We have quarterly appraisal data for 41 commercial office REIT office buildings located in Tokyo for the 44 quarters starting at Q1:2005 and ending at Q4:2015.**
- **The builder's model using appraisal data is somewhat different from the builder's model using selling price data.**
- **The panel nature of the REIT data means that we can use a single property specific dummy variable as a variable that concentrates all of the location attributes of the property into a single variable.**
- **There are 41 separate properties in our REIT data set. For each of our 44 quarters, we assume that the 41 properties appear in the appraised property value for property n in period t, V_{tn}, in the same order.**
- **Our REIT model using appraisal data is on the next slide.**

The Builder's Model Using Property Appraisal Data (cont)

- ω_n in (25) below now the ***property n sample average land price*** (per m²) rather than a Ward n relative price of land:

(25) $V_{tn} = \sum_{n=1}^{41} \omega_n L_{tn} + p_{St}(1 - 0.025)^{A(t,n)} S_{tn} + \varepsilon_{tn}$.

- Thus in **Model 1** above, there are no quarter t land price parameters in this very simple model with 41 unknown property average land price ω_n parameters to estimate.
- Note that the geometric (net) depreciation rate in the model defined by (25) was assumed to be 2.5% per year.
- The final log likelihood for this model was –14968.77 and the R^2 was **0.9426**.
- Thus this very simple model explains most of the variation in the data.
- In our next model, we introduce time dummy variables for the land prices. (Why did we not do this in Model 1 instead of introducing property dummy variables?)

The Builder's Model Using Appraisal Data: Model 2

- **In Model 2, we introduce quarterly land prices α_t into the above model. The new nonlinear regression model is the following one:**

(26) $V_{tn} = \sum_{n=1}^{41} \alpha_t \omega_n L_{tn} + p_{St}(1 - 0.025)^{A(t,n)} S_{tn} + \varepsilon_{tn}$

- **Not all of the quarterly land price parameters (the α_t) and the average property price parameters (the ω_n) can be identified. Thus we impose the following normalization on our coefficients:**

(27) $\alpha_1 = 1$.

- **We used the final parameter values for the ω_n from Model 1 as starting coefficient values for Model 2 (with all α_t initially set equal to 1).**
- **The final log likelihood for Model 2 was –13999.00, a huge improvement of 969.77 for adding 43 new parameters.**
- **The R^2 was 0.9804. Thus the 41 property average price parameters ω_n and the 43 quarterly average land price parameters α_t explain most of the variation in the data.**

The Builder's Model Using Appraisal Data: Model 3

- **Model 3 is the following nonlinear regression model:**

(28) $V_{tn} = \alpha_t \omega_n L_{tn} + p_{St}(1 - \delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}$;

- **where δ is the annual geometric (net) depreciation rate.**
- **The normalization (27) is also imposed.**
- **Thus Model 3 is the same as Model 2 except that we now estimate the single geometric depreciation rate δ.**
- **We used the final parameter values for the α_t and ω_n from Model 2 as starting coefficient values for Model 3 (with δ initially set equal to 0.025).**
- **The final log likelihood for this model was –13993.47, and increase of 5.53 for one additional parameter, and the R^2 was 0.9806.**
- **The sequence of land price (per m^2) α_t, for t =1,2,...,44 is our estimated sequence of quarterly Tokyo land prices, $PL_{REIT}{}^t$, which appears in Chart 3 below.**

The Builder's Model Using Appraisal Data: Model 3 (cont)

- **The estimated geometric (net) depreciation rate was δ = 0.01353.**
- **We also estimated the straight line depreciation model counterpart to Model 3.**
- **The resulting estimated straight line depreciation rate was equal to 0.01317 (t statistic = 45.73).**
- **The R^2 for this model was 0.9806 and the final log likelihood was -13989.83. (pretty close to –13993.47)**
- **The resulting land price series was very similar to the land price series generated by Model 3 above.**
- **Recall that our estimated REIT Model 3 geometric depreciation rate δ was only 1.35% per year which is much lower than our estimated MLIT single geometric depreciation rate from Model 7 above which was 3.41% per year.**
- **Thus the appraisal data and the sales transaction data generate very different geometric depreciation rates.**

The Builder's Model Using Tax Assessment Data

- **We used the Official Land Price (OLP) data described in section 2 above.**
- **We have 6242 annual assessed values for the land components of commercial properties in Tokyo covering the 11 years 2005-2015. We will label these years as t = 1,2,...,11. The assessed land value for property n in year t is denoted as V_{tn}.**
- **We have information on which Ward each property is located and the ward dummy variables $D_{W,tnj}$ are defined by definitions (4) above.**
- **The land plot area of property n in year t is denoted by L_{tn} and the subway variables DS_{tn} and TT_{tn} are defined as in section 2 above.**
- **The number of observations in year t is N(t).**

The Builder's Model Using Tax Assessment Data: Model 1

- **Our initial regression model is the following one where we regress property land value on the ward dummy variables times the land plot area:**

(29) $V_{tn} = (\sum_{j=1}^{23} \omega_j D_{W,tnj})L_{tn} + \varepsilon_{tn}$

- **Thus in Model 1 above, there are no year t land price parameters in this very simple model and ω_j is an estimate of the average land price (per m^2) in Ward j for j = 1,...,23.**
- **The final log likelihood for this model was –67073.91 and the R^2 was 0.3647.**
- **Since we no longer have panel data, the R^2 will be much lower than the R^2 we obtained when we used appraisal data.**
- **In the next model, we will introduce time dummy variables that will lead to our land price index.**

The Builder's Model Using Tax Assessment Data: Model 2

- In Model 2, we introduce annual land prices α_t into the above model. The new nonlinear regression model is the following one:

(30) $V_{tn} = \alpha_t(\sum_{j=1}^{23} \omega_j D_{W,tnj})L_{tn} + \varepsilon_{tn}$;

- Not all of the 11 annual land price parameters (the α_t) and the 23 Ward average property relative price parameters (the ω_n) can be identified.
- Thus we impose the normalization $\alpha_1 = 1$.
- We used the final parameter values for the ω_n from Model 1 as starting coefficient values for Model 2 (with all α_t initially set equal to 1).
- The final log likelihood for Model 2 was –67022.90, an increase of 51.01 for adding 43 new parameters.
- The R^2 was 0.3748. (Still pretty low!)

The Builder's Model Using Tax Assessment Data: Model 3

- In our next model, we allowed the price of land to vary as the lot size increased. We divided up our 6242 observations into 5 groups of observations based on their lot size.
- We define the *5 land dummy variables*, $D_{L,tnk}$, for k = 1,...,5 as follows:

(31) $D_{L,tnk} \equiv 1$ if observation tn has land area that belongs to group k;

$\equiv 0$ if observation tn has land area that does not belong to group k.

- Define the constants L_1-L_4 as 100, 150, 200 and 300 respectively.
- These constants and the dummy variables defined by (31) are used in the definition of the following piecewise linear function of L_{tn}, $f(L_{tn})$:

The Builder's Model Using Assessment Data: Model 3 (concl.)

(32) $f(L_{tn}) \equiv D_{L,tn1}\lambda_1 L_{tn} + D_{L,tn2}[\lambda_1 L_1+\lambda_2(L_{tn}-L_1)]$
$+ D_{L,tn3}[\lambda_1 L_1+\lambda_2(L_2-L_1)+\lambda_3(L_{tn}-L_2)]$
$+ D_{L,tn4}[\lambda_1 L_1+\lambda_2(L_2-L_1)+\lambda_3(L_3-L_2)+\lambda_4(L_{tn}-L_3)]$
$+ D_{L,tn5}[\lambda_1 L_1+\lambda_2(L_2-L_1)+\lambda_3(L_3-L_2)\ )+\lambda_4(L_4-L_3)+\lambda_5(L_{tn}-L_4)]$.

- **Model 3** was defined as the following nonlinear regression model:

(33) $V_{tn} = \alpha_t(\sum_{j=1}^{23} \omega_j D_{W,tn,j})f(L_{tn}) + \varepsilon_{tn}$

- We imposed the normalizations $\alpha_1 = 1$ and $\lambda_1 = 1$ so that all of the remaining parameters in (33) could be identified.
- We used the final parameter values for the α_t and ω_j from Model 2 as starting coefficient values for Model 3 (with all λ_k initially set equal to 1). Thus Model 3 adds the 4 new marginal prices of land, λ_2, λ_3, λ_4 and λ_5 to Model 2.
- The final log likelihood for Model 3 was –66044.02, an increase of 978.88 for adding 4 new parameters. (A huge increase).
- The R^2 was 0.4668.
- Our final land price model added the subway variables to Model 3.

The Builder's Model Using Tax Assessment Data: Model 4

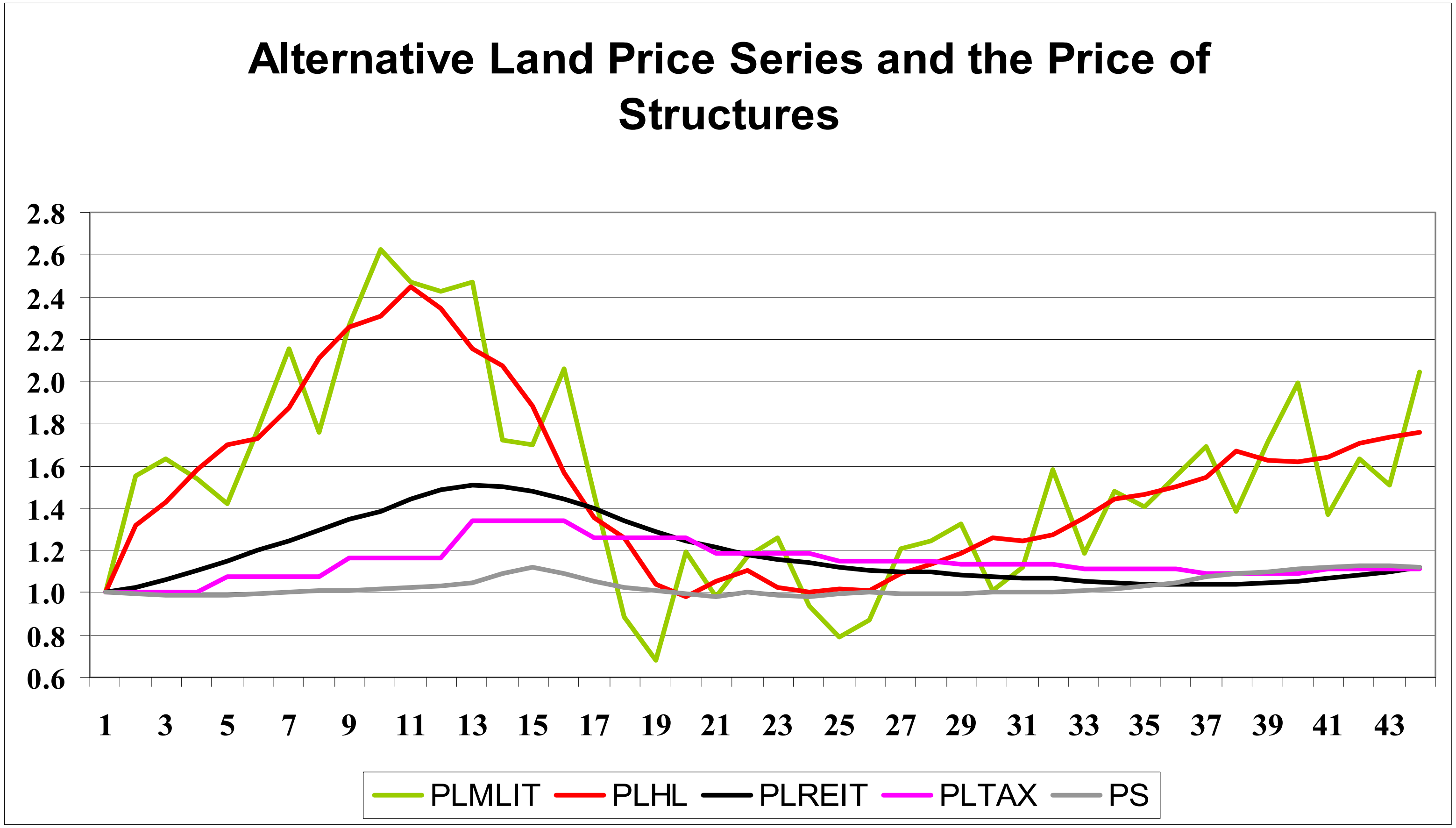
- **Model 4 was defined as the following nonlinear regression model:**

(34) $V_{tn} = \alpha_t(\sum_{j=1}^{23} \omega_j D_{W,tnj})\,(1+\eta(DS_{tn}-50))(1+\theta(TT_{tn}-4))f(L_{tn}) + \varepsilon_{tn}.$

- **Model 4 has added two new subway parameters, η and θ, to Model 3.**
- **The final log likelihood for Model 4 was –65584.56, an increase of 459.46 for adding 2 new parameters.**
- **The R^2 was 0.5401.**
- **The α_t sequence of estimated parameters (along with $\alpha_1 \equiv 1$) forms an annual (quality adjusted) Official Land Price series.**
- **For comparison purposes, we repeat each α_t four times and convert the annual Official Land Price series into the quarterly Official Land Price series, $PL_{OLP}{}^t$.**
- **This land price series is compared with our final transactions based MLIT land price series $PL_{MLIT}{}^t$ and its linear smooth $PL_L{}^t$ along with our final REIT based land price series $PL_{REIT}{}^t$ in the next slide.**

Comparing Land Price Indexes from Different Sources

The green line is the MLIT Land Price Index; the red line is its Henderson linear smooth. The black line is the REIT based Land Price Index and the purple line is the tax assessment based Land Price Index. The grey (almost constant) line is the structures Price Index. We like the linear smooth!



Alternative Overall Commercial Property Price Indexes

- In the property price literature, a frequently used index of overall property prices is the period average of the individual property values V_{tn} divided by the corresponding structure area S_{tn}.
- Thus define the (preliminary) quarter t Mean Property Price Index $P_{MEANP}{}^{t}$ as follows:

(35) $P_{MEANP}{}^{t} \equiv (1/N(t))\Sigma_{n=1}{}^{N(t)} V_{tn}/S_{tn}$;

- The final mean property price index for quarter t, $P_{MEAN}{}^{t}$, is defined as the corresponding preliminary index $P_{MEANP}{}^{t}$ divided by $P_{MEANP}{}^{1}$; i.e., we normalize the series defined by (35) to equal 1 in quarter 1.
- The mean property price series $P_{MEAN}{}^{t}$ is rather volatile and so we smooth it using the Henderson Linear Smoothing Method that we applied to the MLIT Land Price series.
- The resulting Smoothed Mean Property Price Index is denoted by $P_{MEANS}{}^{t}$.

Alternative Overall Commercial Property Price Indexes (cont)

- **We can use the predicted values from the MLIT Model 11 regression in order to construct quarterly estimates for the price and quantity of commercial land and the corresponding price and quantity of constant quality commercial structures.**
- **We then combined these land and structure series into an overall MLIT Chained Fisher Property Price Index which we denote by P_{FMLIT}^{t} for quarter t.**
- **This series is also quite volatile so we used the Henderson type linear smoothing procedure to construct the Smoothed MLIT Fisher Property Price Index P_{FMLITS}^{t}.**
- **We also used the results from Model 3 that used the REIT data to construct quarterly estimates for the price and quantity of commercial land and structures and we combined these estimates into the REIT Based Property Price Index P_{FREIT}^{t} .**
- **This series was not volatile and did not require any smoothing.**

Alternative Overall Commercial Property Price Indexes (cont)

- **Our final property price index will be generated by a traditional log price time dummy hedonic regression using the MLIT data.**
- **We use the same notation and definitions of variables as was used in Section 4 above.**
- **Define the natural logarithms of V_{tn}, L_{tn} and S_{tn} as LV_{tn}, LL_{tn} and LS_{tn} for $t = 1,\ldots,44$ and $n = 1,\ldots,N(t)$.**
- **The log price time dummy hedonic regression model is the following linear regression model:**

$$(42)\quad LV_{tn} = \beta_t + \sum_{j=2}^{4}\omega_j D_{W,tnj} + \gamma A_{tn} + \lambda LL_{tn} + \mu LS_{tn} + \sum_{h=4}^{10} \phi_h D_{H,tnh} + \eta DS_{tn} + \theta TT_{tn} + \varepsilon_{tn} .$$

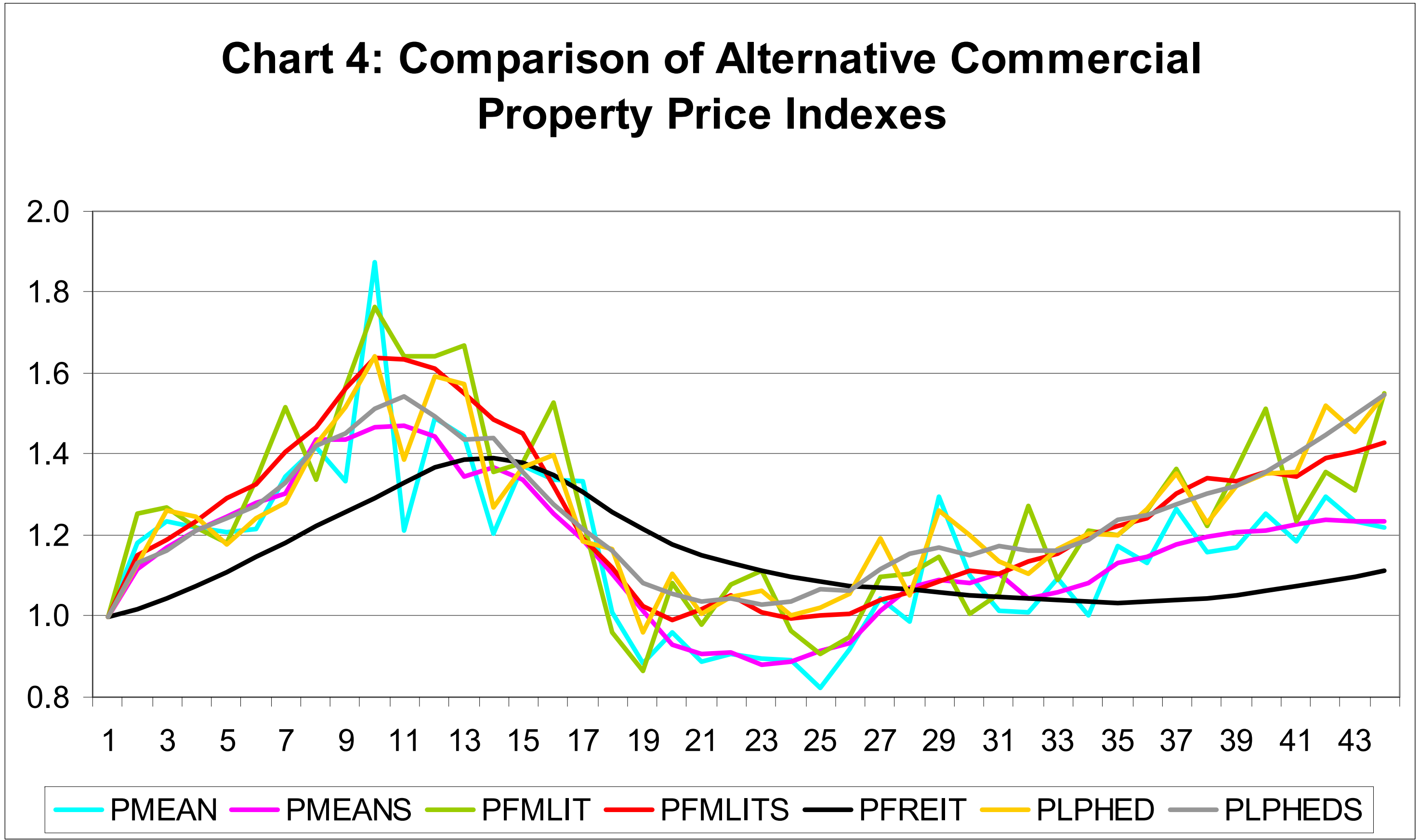
- **The R^2 for this regression was 0.7593. This is higher than our Model 9 and Model 11 R^2.**

The Traditional Log Price Time Dummy Hedonic Model

- **Define the unnormalized land price for quarter t, α_t, as the exponential of β_t; i.e., $\alpha_t \equiv \exp(\beta_t)$ for t = 1,...,44.**
- **The log price hedonic regression property price for quarter t, P_{LPHED}^t is defined as α_t/α_1 for t = 1,...,44.**
- **This traditional Hedonic Regression Model Property Price Index is denoted by P_{LPHED}^t in the Chart which follows.**
- **It is possible to convert the estimated age coefficient γ into an estimate for a geometric rate of structure depreciation, δ. The formula for this conversion is $\delta \equiv 1 - e^{\gamma/\beta}$.**
- **The implied δ is 0.01945; i.e., the traditional hedonic regression model generates an implied annual geometric depreciation rate equal to 1.945% per year, which is a reasonable estimate.**
- **The time dummy hedonic regression model property price index P_{LPHED}^t is also too volatile so we applied our modified Henderson linear smoothing operator to P_{LPHED}^t which produced the smoothed series, P_{LPHEDS}^t.**

Comparison of Alternative Property Price Indexes

The 3 jagged lines are P_{MEAN}^{t}, P_{FMLIT}^{t}, and P_{LPHED}^{t}. Their linear smooths are P_{MEANS}^{t}, P_{FMLITS}^{t} and P_{LPHEDS}^{t}. P_{MEAN}^{t} and P_{MEANS}^{t} are too low because they omit depreciation. P_{FREIT}^{t} is too smooth and its turning points lag too much. We like P_{FMLITS}^{t} and P_{LPHEDS}^{t}.



Conclusions

- **It is possible to construct a quarterly transactions based commercial property price index that can be decomposed into *land* and *structure* components.**
- **The main characteristics of the properties that are required in order to implement our approach are: (i) the property location (or neighbourhood); (ii) the floor space area of the structure on the property; (iii) the area of the land plot; (iv) the age of the structure and (v) the height of the building. We also require an appropriate exogenous commercial property construction price index.**
- **The land price index that our hedonic regression model generates may be too volatile and hence may need to be smoothed. We found that a slightly modified five quarter moving average of the raw land price indexes did an adequate job of smoothing. This means that the final land price index could be produced with a two quarter lag.**

Conclusions (continued)

- **We found that a smoothed version of a traditional log price time dummy hedonic regression model produced an acceptable approximation to our preferred smoothed builder's model overall price index.**
- **We also found that a very simple overall price index which is proportional to the quarterly arithmetic average of each property price divided by the corresponding structure area provided a rough approximation to our preferred price index. This model cannot take depreciation into account and hence will in general have an downward bias but it has the advantage of requiring information on only a single property characteristic (the structure floor space area) in order to be implemented.**
- **The price indexes that were based on appraisal and assessed value information were not satisfactory approximations to the transactions based indexes. The turning points in these series lagged our preferred series and the appraisal based series smoothed the data based series to an unacceptable degree.**

Conclusions (concluded)

- The two versions of the builder's model that estimated multiple (net) depreciation rates produced virtually the same indexes and virtually identical depreciation schedules. These rates of depreciation changed materially as the structure aged and the depreciation rates became appreciation rates for structures over age 40.
- Our overall conclusion is that it should be possible for national income accountants to construct acceptable commercial land price series using transactions data on the sales of commercial properties. The required information on the characteristics of the properties is being collected by some private sector businesses. It should be possible for government statisticians to collect the same information using building permit, land registry and property assessment data.

Chihiro Shimizu(清水千弘), PhD

Professor, Nihon University,
34-1, Shimouma 3 chome, Setagaya, Tokyo,154-8513
Tel：(+81)-(0)3-6453-1715（Office）, Fax:03-6453-1630
e-mail: shimizu.chihiro@nihon-u.ac.jp

Research Affiliate, Center for Real Estate
Massachusetts Institute of Technology
105 Massachusetts Ave, Cambridge, MA 02139
Tel: (+1) 617-253-4373,(Office), Fax: (+1) 617-253-4373
e-mail: cshimizu@mit.edu