Residential Property Price Indexes for Tokyo

Erwin Diewert and Chihiro Shimizu[1] April 10, 2013

Discussion Paper 13-07,
School of Economics,
The University of British Columbia,
Vancouver, Canada, V6T 1Z1.

Abstract

The paper uses hedonic regression techniques in order to decompose the price of a house into land and structure components using real estate sales data for Tokyo. In order to get sensible results, a nonlinear regression model using data that covered multiple time periods was used. Collinearity between the amount of land and structure in each residential property leads to inaccurate estimates for the land and structure value of a property. This collinearity problem was solved by using exogenous information on the rate of growth of construction costs in Tokyo in order to get useful constant quality subindexes for the price of land and structures separately.

Key Words

House price indexes, land and structure components, time dummy hedonic regressions, spline functions, flexible functional forms, Fisher ideal indexes.

Journal of Economic Literature Classification Numbers

C2, C23, C43, D12, E31, R21.

---

[1] W. Erwin Diewert: School of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and the School of Economics, University of New South Wales, Sydney, Australia (email: erwin.diewert@ubc.ca) and Chihiro Shimizu, Reitaku University, Kashiwa, Chiba, 277-8686, Japan and the School of Economics, University of British Columbia, (email: cshimizu@reitaku-u.ac.jp). The first author gratefully acknowledges the financial support of the SSHRC of Canada.

1. Introduction

In this paper, we will use hedonic regression techniques in order to construct a quarterly constant quality price index for the sales of residential properties in Tokyo for the years 2000-2010 (44 quarters in all). The usual application of a time dummy hedonic regression model to sales of houses does not lead to a decomposition of the sale price into a structure component and a land component. But such a decomposition is required for many purposes. Our paper will attempt to use hedonic regression techniques in order to provide such a decomposition for Tokyo house prices. Instead of entering characteristics into our regressions in a linear fashion, we enter them as piece-wise linear functions or spline functions to achieve greater flexibility.

The Tokyo house price data that we use will be described in section 2.

In section 3, we will outline our basic (nonlinear) regression model which requires information on the selling price of the property V along with the following *basic characteristics* of the property:

- The land area of the property (L);
- The livable floor space area of the structure (S);
- The age of the structure (A) and
- The location of the property.

Using only information on these 4 characteristics plus the use of an exogenous residential house construction price index for Tokyo, we are able to explain 0.8168 percent of the variation in the sales data. Our basic nonlinear regression model is a variant of the *builder's hedonic regression model* introduced by Diewert, de Haan and Hendriks (2011a)(2011b).

In section 4, we introduced some additional parameters into the model without requiring additional information on characteristics. Instead of assuming a single straight line depreciation rate for the structure, we allowed the depreciation rate to follow a piecewise linear structure. We also allowed the price of land per square meter for a property to follow a piecewise linear structure. For the addition of 4 parameters over the model in section 3, the R^2 of our model increased from 0.8168 to 0.8206 and the log likelihood increased by 68.9.

In sections 5 and 6, we used information on some additional characteristics of the properties sold in each quarter. In section 5, we utilized information on the *number of bedrooms* NB and the *width of the lot* W, adding an additional 6 parameters to our nonlinear regression model. The R^2 of our new model increased from 0.8206 to 0.8256 and the log likelihood increased by 78.7. In section 6, we utilized information on the *time it takes to walk to the nearest subway* TW and the *time it takes to go from the nearest* *subway station to downtown Tokyo* TT, adding an additional 6 parameters to our regression model. The R^2 of our new model increased from 0.8256 for the section 5 model to 0.8417 and the log likelihood increased by a very large 269.4.

In section 7, we divided the 22 wards in Tokyo that appear in our regression models into *expensive wards* and *inexpensive wards* and we allow the movements in the price of land to be different in these two classes of wards. This generalization of our earlier models added 45 parameters to be estimated. The R^2 of our new model increased from 0.8417 for the section 6 model to 0.8476 and the log likelihood increased by 106.0. At this point, we stopped adding additional characteristics to our model and judged the section 7 model to be satisfactory.

In section 8, we switch our attention from compiling land, structure and overall house price indexes for *sales* of residential properties to the problems associated with constructing the corresponding indexes for the *stock* of residential housing in Tokyo. We did not have access to information on the total stock of residential houses in Tokyo over time but we used the total number of houses transacted over our sample period as an approximation to the total stock. The resulting approximate stock prices for selected models are listed in this section.

In section 9, we take the model explained in section 7 but estimate the parameters over a 5 year *rolling window* period. We use the estimated indexes for the last two periods in each rolling window regression to update our previous index. The resulting index is meant to approximate a realistic house price index that could be implemented by a statistical agency. We find little differences between the resulting Rolling Window estimates and the estimates obtained in section 7.[2]

In section 10, we compare our section 7 overall house price indexes that were constructed using our nonlinear hedonic regression with two *typical time dummy hedonic regression* that uses the log of selling prices as the dependent variable. This typical hedonic regression approach cannot be used to generate realistic prices of land and structures but the overall house price index generated by this typical approach can be compared with our overall house price index. We find that the general pattern between the three overall indexes is much the same but our section 7 time dummy index generates higher prices than the corresponding indexes generated by the time dummy approach.

Section 11 concludes.

2. The Tokyo Housing Data

Our basic data set on V, L, S, A, the location of the property and some additional characteristics to be explained below was obtained from a weekly magazine, *Shukan Jutaku Joho* (Residential Information Weekly) published by Recruit Co., Ltd., one of the

[2] Rolling Window time dummy hedonic regressions were used by Shimizu, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010). A special case of the Rolling Window methodology is the adjacent year time dummy hedonic regression introduced by Court (1939; 109-111).

largest vendors of residential listings information in Japan. The Recruit dataset covers the 23 special wards of Tokyo for the period 2000 to 2010, including the mini-bubble period in the middle of 2000s and its later collapse caused by the Great Recession. *Shukan Jutaku Joho* provides time series of housing prices from the week when it is first posted until the week it is removed due to its sale.[3] We only use the price in the final week because this can be safely regarded as sufficiently close to the contract price.[4]

There were a total of 5578 observations (after range deletions) in our sample of sales of single family houses in the Tokyo area over the 44 quarters covering 2000-2010.[5] The definitions for the above variables and their units are as follows:

V = The value of the sale of the house in 10,000,000 Yen;
S = Structure area (floor space area) in units of 100 meters squared;
L = Lot area in units of 100 meters squared;
A = Approximate age of the structure in years;
NB = Number of bedrooms;
WI = Width of the lot in meters;
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).

The basic descriptive statistics for the above variables are listed in Table 1 below.

Table 1: Descriptive Statistics for the Variables

Name	No. of Obs.	Mean	Std. Dev	Minimum	Maximum
V	5578	6.2310	2.95420	2.0500	20
S	5578	1.0961	0.36255	0.5012	2.4789
L	5578	1.0283	0.42538	0.5001	2.4977
A	5578	14.689	8.91460	2.0140	49.7230
NB	5578	3.9518	1.04090	2	8
WI	5578	4.6987	1.26090	2.5	9
TW	5578	9.9295	4.48510	2	29
TT	5578	31.677	7.55220	4	48

Thus over the sample period, the sample average sale price was approximately 62.3 million Yen, the average structure space was 110 m^2, the average lot size was 103 m^2, the average age of the structure was 14.7 years, the average number of bedrooms in the

[3] There are two reasons for the listing of a unit being removed from the magazine: a successful deal or a withdrawal (i.e. the seller gives up looking for a buyer and thus withdraws the listing). We were allowed access to information regarding which the two reasons applied for individual cases and we discarded those transactions where the seller withdrew the listing.

[4] Recruit Co., Ltd. provided us with information on contract prices for about 24 percent of all listings. Using this information, we were able to confirm that prices in the final week were almost always identical with the contract prices; see Shimizu, Nishimura and Watanabe (2012).

[5] We deleted 9.2 per cent of the observations because they fell outside our range limits for the variables V, L, S, A, NB and W. It is risky to estimate hedonic regression models over wide ranges when observations are sparse at the beginning and end of the range of each variable. The a priori range limits for these variables were as follows: $2 \leq V \leq 20$; $0.5 \leq S \leq 2.5$; $0.5 \leq V \leq 2.5$; $1 \leq A \leq 50$; ; $2 \leq NB \leq 8$; $2.5 \leq W \leq 9$.

houses that were sold was 3.95, the average lot width was 4.7 meters, the average walking time to the nearest subway station was 9.9 minutes and the average subway travelling time from the nearest station to the Tokyo Central station was 31.7 minutes.

There were fairly high correlations between the V, S and L variables. The correlations of the selling price V with structure and lot area S and L were 0.689 and 0.660 respectively and the correlation between S and L was 0.668. Given the large amount of variability in the data and the relatively high correlations between V, S and L, we can expect multicollinearity problems in a simple linear regression of V on S and L.[6]

In order to eliminate the multicollinearity problem between the lot size L and floor space area S for an individual house, we will assume that the value of a new structure in any quarter is proportional to a Construction Cost Price Index for Tokyo.[7]

In addition to having the information listed in Table 1 on residential houses sold in Tokyo over 2000-2010, we also had the address for each transaction. We used this information in order to allocate each sale into one of 21 Wards for the Tokyo area. We constructed Ward dummy variables and made use of these variables in most of our regressions as locational explanatory variables.

3. The Basic Builder's Model with Locational Dummy Variables

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.

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 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. Now think of a sample of properties of the same general type, which have prices or values V_{tn} in period t[8] and structure areas S_{tn} and land areas L_{tn} for n = 1,...,N(t) where N(t) is the number of observations in period t. 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*

[6] See Diewert, de Haan and Hendriks (2011a) (2011b) for evidence on this multicollinearity problem using Dutch data.

[7] This index was constructed by the Construction Price Research Association which is now an independent agency but prior to 2012 was part of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), a ministry of the Government of Japan. The quarterly values for this index are listed in Table A2 in the Appendix; see the listing for the variable P_{S1}. The quarterly values were constructed from the Monthly Residential Construction Cost index for Tokyo.

[8] The period index t runs from 1 to 44 where period 1 corresponds to Q1 of 2000 and period 44 corresponds to Q4 of 2010.

regression model for period t where the α_t and β_t are the parameters to be estimated in the regression:[9]

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

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. Finally, note that separate linear regressions can be run of the form (1) for each period t in our sample.

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_{tn}, and assuming a straight line depreciation model, a more realistic hedonic regression model than that defined by (1) above is the following *basic builder's model*:[10]

(2) $V_{tn} = \alpha_t L_{tn} + \beta_t(1 - \delta_t A_{tn})S_{tn} + \varepsilon_{tn}$; $t = 1,...,44; n = 1,...,N(t)$

where the parameter δ_t reflects the *net 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 0.25 and 2.5%.[11] Note that (2) is now a nonlinear regression model whereas (1) was a simple linear regression model. Both models (1) and (2) can be run period by period; it is not necessary to run one big regression covering all time periods in the data sample. The period t price of land will the estimated coefficient for the parameter α_t and the price of a unit of a newly built structure for period t will be the estimate for β_t. The period t quantity of land for property n is L_{tn} and the period t quantity of structure for property n, expressed in equivalent units of a new structure, is $(1 - \delta_t A_{tn})S_{tn}$ where S_{tn} is the floor space area of property n in period t.

Note that the above model is a *supply side model* as opposed to the *demand side model* of Muth (1971) and McMillen (2003). Basically, we are assuming competitive suppliers of

[9] Other papers that have suggested hedonic regression models that lead to additive decompositions of property values into land and structure components include Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic, Longhofer and Redfearn (2007), Davis and Heathcote (2007), Francke (2008), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert (2010) (2011), Rambaldi, McAllister, Collins and Fletcher (2010) and Diewert, Haan and Hendriks (2011a) (2011b).

[10] This formulation follows that of Diewert (2010) (2011) and Diewert, Haan and Hendriks (2011a) (2011b). It is a special case of Clapp's (1980; 258) hedonic regression model.

[11] This estimate of depreciation is regarded as a *net depreciation rate* because it is equal to a "true" gross structure depreciation rate less an average renovations appreciation rate. Since we do not have information on renovations and additions to a structure, our age variable will only pick up average gross depreciation less average real renovation expenditures. Note that we excluded sales of houses from our sample if the age of the structure exceeded 50 years when sold. Very old houses tend to have larger than normal renovation expenditures and thus their inclusion can bias the estimates of the net depreciation rate for younger structures.

housing so that we are in Rosen's (1974; 44) Case (a), where the hedonic surface identifies the structure of supply. This assumption is justified for the case of newly built houses but it is less well justified for sales of existing homes.[12]

As was mentioned in the previous section, we have 5578 observations on sales of houses in Tokyo over the 44 quarters in years 2000-2010. Thus equations (2) above could be combined into one big regression and a single depreciation rate $\delta = \delta_t$ could be estimated along with 44 land prices α_t and 44 new structure prices β_t so that 89 parameters would have to be estimated. However, 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.[13] Thus in order to deal with the multicollinearity problem, we draw on *exogenous information* on new house building costs from the Japanese Ministry of Land, Infrastructure, Transport and Tourism (MLIT) and we assume that the price of new structures is proportional to this index of residential building costs. Thus our new builder's model that uses exogenous information on structure prices is the following one:

$$(3)\ V_{tn} = \alpha_t L_{tn} + \beta p_{Ct}(1 - \delta A_{tn})S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,\ldots,44;\ n = 1,\ldots,N(t)$$

where all variables have been defined above except that p_{Ct} is the MLIT house construction cost index for Tokyo for quarter t. Thus we have 5578 degrees of freedom to estimate 44 land price parameters α_t, one structure price parameter β that determines the level of prices over our sample period and one annual straight line depreciation rate parameter δ, a total of 46 parameters.

The R^2 for the resulting nonlinear regression model was only 0.5704,[14] which is not very satisfactory. Thus the simple Builder's Model defined by (3) was not as satisfactory as was the corresponding Builder's Model for the small town of "A" in the Netherlands where the R^2 was 0.8703 using the same information on characteristics of the house and lot. However, in the case of the town of "A", the structures were all much the same and all houses in the town had access to basically the same amenities. The situation in the huge city of Tokyo is very different: different neighborhoods have access to very different amenities and Tokyo is not situated on a flat, featureless plain and so we would expect substantial variations in the price of land across the various neighborhoods.

---

[12] Thorsnes (1997; 101) assumed that a related supply side model held instead of equation (2). He assumed that housing was produced by a CES production function $H(L,K) \equiv [\alpha L^{\rho}+\beta K^{\rho}]^{1/\rho}$ where K is structure quantity and $\rho \neq 0$; $\alpha > 0$; $\beta > 0$ and $\alpha+\beta = 1$. He assumed that property value V_n^t is equal to $p_t H(L_n^t,K_n^t)$ where p_t, ρ, α and β are parameters to be estimated. However, our builder's model assumes that the production functions that produce structure space and that produce land are independent of each other.

[13] See Schwann (1998) and Diewert, de Haan and Hendriks (2011a) and (2011b) on the multicollinearity problem.

[14] All of the R^2 reported in this paper are equal to the square of the correlation coefficient between the dependent variable in the regression and the corresponding predicted variable. The estimated net annual straight line depreciation rate was $\delta = 1.25\%$, with a T statistic of 17.3. Due to the poor fit of the model, we will not report the other estimated parameters.

In order to take into account possible neighbourhood effects on the price of land, we introduced *ward dummy variables*, $D_{W,tn,j}$, into the hedonic regression (3). These 21 dummy variables are defined as follows: for t = 1,...,44; n = 1,...,N(t); j = 1,...,21:[15]

(4) $D_{W,tn,j} \equiv 1$ 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 (3) to allow the *level* of land prices to differ across the 21 Wards of Tokyo. The new nonlinear regression model is the following one:

(5) $V_{tn} = \alpha_t(\sum_{j=1}^{21} \omega_j D_{W,tn,j})L_{tn} + \beta p_{Ct}(1 - \delta A_{tn})S_{tn} + \varepsilon_{tn}$; $\quad$ t = 1,...,44; n = 1,...,N(t).

Comparing the models defined by equations (3) and (5), it can be seen that we have added an additional 21 *ward relative land value parameters*, $\omega_1,...,\omega_{21}$, to the model defined by (3). However, looking at (5), it can be seen that the 44 land time parameters (the α_t) and the 21 ward parameters (the ω_j) cannot all be identified. Thus we need to impose at least one identifying normalization on these parameters. We chose the following normalization:

(6) $\omega_{10} \equiv 1$.

We will call the hedonic regression model defined by (5) and (6) *Model 1*. The tenth ward, Setagay, has the most transactions in our sample (1158 transactions over the sample period) and thus the level of land prices in this Ward should be fairly accurately determined. Hence the remaining ω_j represent the level of land prices in Ward j *relative* to the level in Ward 10 so if say $\omega_1 > 1$, this means that on average, the price of land in Ward 1 is higher than the average price of land in Ward 10. Taking into account the normalization (6), it can be seen that Model 1 has 44 unknown land price parameters α_t, 20 ward relative land price parameters ω_j, one structure price level parameter β and one annual net depreciation parameter δ that need to be estimated. We estimated these parameters using the nonlinear regression option in Shazam; see White (2004). The detailed parameter estimates are listed in the Appendix in Table A1.[16] The R^2 for this model turned out to be 0.8168 and the log likelihood (LL) was −9233.0, a huge increase of 2270.6 over the LL of the model defined by (3). Thus the Ward variables are very significant determinants of Tokyo house prices.

---

[15] The 21 Wards of Tokyo that had at least one transaction during our sample period (with the total number of transactions for that Ward in brackets) are as follows: 1: Minato (69); 2: Shinjuku (136); 3: Bunkyo (82); 4: Taito (15); 5: Sumida (32); 6: Koto (38); 7: Shinagawa (144); 8: Meguro (349); 9: Ota (409); 10: Setagay (1158); 11: Shibuya (107); 12: Nakano (305); 13: Suginami (773); 14: Toshima (124); 15: Kita (53); 16: Arakawa (34); 17: Itabashi (214); 18: Nerima (925); 19: Adachi (271); 20: Katsushika (143); 21: Edogawa (197). Note that for each observation tn, we have $\sum_{j=1}^{21} DW_{tn,j} = 1$; i.e., for each observation tn, the 21 ward dummy variables sum to one. Recall that there are 5578 observations in our sample.

[16] We note that the annual net depreciation rate for Model 1 was estimated as $\delta = 1.39\%$ with a T statistic of 26.8.

We regard Model 1 as a minimally satisfactory model. Note that we used only four characteristics for each house sale: the land area L, the structure area S, the age of the structure A and its Ward location.

We now address the problem of how exactly should the land, structure and overall house price index be constructed? Our nonlinear regression model defined by (5) decomposes into two terms: one which involves the land area L_{tn} of the house, $\alpha_t(\sum_{j=1}^{21} \omega_j D_{W,tn,j})L_{tn}$, and another which involves the structure area S_{tn} of the house, $\beta p_{Ct}(1 - \delta A_{tn})S_{tn}$. The first term can be regarded as an estimate of the land value of house n that was sold in quarter t while the second term is an estimate of the structure value of the house. Our problem now is how exactly should these two value terms be decomposed into constant quality price and quantity components? Our view is that a suitable constant quality land price index for all houses sold in period t should be α_t and for house n sold in period t, the corresponding constant quality quantity should be $(\sum_{j=1}^{21} \omega_j D_{W,tn,j})L_{tn}$ which in turn is equal to $\omega_j L_{tn}$ if house n sold in period t is in Ward j.[17] The basic idea here is that we regard the term $\alpha_t(\sum_{j=1}^{21} \omega_j D_{W,tn,j})L_{tn}$ as a time dummy hedonic model for the land component of the house with α_t acting as the time dummy coefficient. Thus if we priced out house n that sold in period t in period s, our hedonic imputation[18] for the land component of this "model" would be $\alpha_s(\sum_{j=1}^{21} \omega_j D_{W,tn,j})L_{tn}$. Thus the quarterly time coefficients α_t act as *proportional time shifters of the hedonic surface for the land component* of the value of each house in our sample and the relative period t to period s land price for each house is α_t/α_s.

Similarly, a suitable constant quality structure price index for all houses sold in period t is βp_{Ct} and for house n sold in period t, the corresponding constant quality quantity should be approximately equal to the depreciated structure quantity $(1-\delta A_{tn})S_{tn}$. Thus we regard the term $\beta p_{Ct}(1-\delta A_{tn})S_{tn}$ as a time dummy hedonic model for the structure component of the house with βp_{Ct} acting as the time dummy coefficient. The quarterly time coefficients βp_{Ct} (or just the p_{Ct}) act as proportional time shifters of the hedonic surface for the structure component of each house in our sample and the period t to period s land price for each house in our sample turns out to be p_{Ct}/p_{Cs}.[19]

---

[17] An alternative way of viewing our land model is that land in each Ward can be regarded as a distinct commodity with its own price and quantity. But since all Ward land prices move proportionally over time, virtually all index number formulae will generate an overall land price series that is proportional to the α_t.

[18] Hedonic imputation models and time dummy hedonic models are discussed in more detail in Diewert (2003b), de Haan (2003), (2008) (2009), Diewert, Heravi and Silver (2009) and de Haan and Diewert (2011).

[19] Our method for aggregating over different house "models" that have varying amounts of constant quality land and structures can be viewed as a hedonic imputation method but it can also be viewed as an application of Hicks' Aggregation Theorem; i.e., if the prices in a group of commodities vary in strict proportion over time, then the factor of proportionality can be taken as the price of the group and the deflated group expenditures will obey the usual properties of a microeconomic commodity. "Thus we have demonstrated mathematically the very important principle, used extensively in the text, that if the prices of a group of goods change in the same proportion, that group of goods behaves just as if it were a single commodity." J.R. Hicks (1946; 312-313).

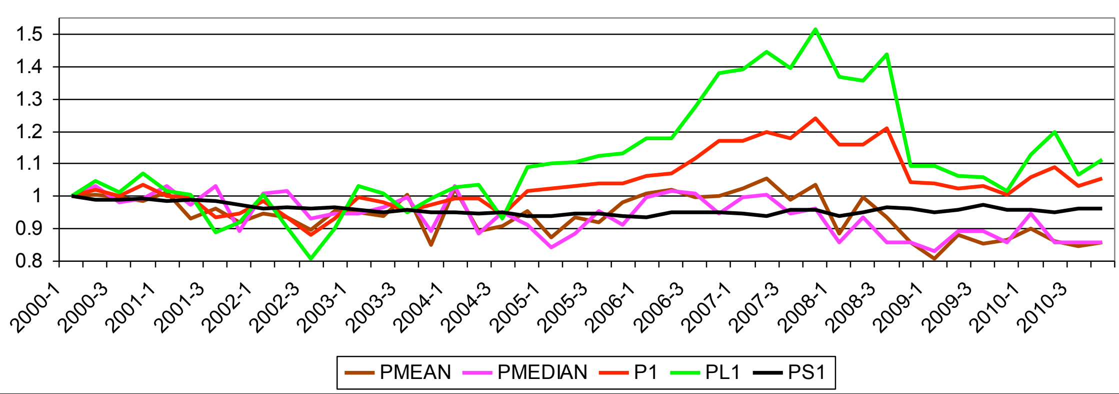
Thus the *constant quality residential land price index* for Tokyo for quarter t is defined to be $P_{L1t} \equiv \alpha_t/\alpha_1$ and the corresponding *constant quality residential structures price index* for Tokyo for quarter t is defined to be $P_{S1t} \equiv p_{Ct}/p_{C1}$.[20] These price indexes can be regarded as quarter t price levels for land and structures respectively and the corresponding Model 1 quarter t *constant quality quantity levels*, Q_{L1t} and Q_{S1t}, are defined as the total quarter t values of land and structures divided by the corresponding price levels for t = 1,...,44:

$$(7)\ Q_{L1t} \equiv \sum_{n=1}^{N(t)} (\sum_{j=1}^{21} \omega_j D_{W,tn,j})\alpha_t L_{tn}/P_{L1t} = \alpha_1 \sum_{n=1}^{N(t)} (\sum_{j=1}^{21} \omega_j D_{W,tn,j}) L_{tn}\ ;$$
$$(8)\ Q_{S1t} \equiv \sum_{n=1}^{N(t)} \beta p_{Ct}(1 - \delta A_{tn})S_{tn}/P_{S1t} = \beta \sum_{n=1}^{N(t)} (1 - \delta A_{tn})S_{tn}\ .$$

The price and quantity series for land and structures need to be aggregated into an overall Tokyo house price index. We use the Fisher (1922) ideal index to perform this aggregation. Thus define the *overall house price level for quarter t* for Model 1, P_{1t}, as the chained Fisher price index of the land and structure series $\{P_{L1t},P_{S1t},Q_{L1t},Q_{S1t}\}$.[21]

The overall Model 1 house price index P_{1t} as well as the land and structure price indexes P_{L1t} and P_{S1t} for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 1 below. We have also computed the quarterly mean and median house prices transacted in each quarter and then normalized these averages to start at 1 in Quarter 1 of 2000. These overall average price index series, P_{Mean} and P_{Median} are also graphed in Chart 1.[22]

Chart 1: Mean, Median and Overall Price, Land Price and Structure Price Indexes for Model 1



[20] We have normalized the price indexes P_{L1t} and P_{S1t} to equal 1 in quarter 1, which is quarter 1 of the year 2000.

[21] The Fisher chained index P_{1t} is defined as follows. For t = 1, define $P_{1t} \equiv 1$. For t > 1, define P_{1t} in terms of P_{1t-1} and P_{Ft} as $P_{1t} \equiv P_{1t-1}P_{Ft}$ where P_{Ft} is the quarter t Fisher chain link index. The chain link index for t ≥ 2 is defined as $P_{Ft} \equiv [P_{Lt}P_{Pt}]^{1/2}$ where the Laspeyres and Paasche chain link indexes are defined as $P_{Lt} \equiv [P_{L1t}Q_{L1t-1}+P_{L1t}Q_{L1t-1}]/[P_{L1t-1}Q_{L1t-1}+P_{L1t-1}Q_{L1t-1}]$ and $P_{Pt} \equiv [P_{L1t}Q_{L1t}+P_{L1t}Q_{L1t}]/[P_{L1t-1}Q_{L1t}+P_{L1t-1}Q_{L1t}]$. Diewert (1976) (1992) showed that the Fisher formula had good justifications from both the perspectives of the economic and axiomatic approaches to index number theory.

[22] The series P_{Mean}, P_{Median}, P_1, P_{L1} and P_{S1} are also listed in Table A2 of the Appendix.

The land price series P_{L1} is the top line in Chart 1, followed by the overall Model 1 house price index P_1, followed by the structure price index P_{S1} (at the end of the sample period). The mean and median price series track each other and our overall price series P_1 reasonably well until 2004 but in the following years, the mean and median series fall well below our overall quality adjusted house price series P_1.[23] Thus quality adjusting the sales of residential housing in Tokyo makes a big difference to the resulting index.

In the following section, we will use our information on lot size and the age of the house in a more flexible regression model and construct the resulting quality adjusted price indexes and compare them with the Model 1 indexes.

4. The Use of Splines on Lot Size and on the Age of the Structure

In most countries, the price of a residential lot as a function of lot size does not grow in a linear fashion as is predicted by our Model 1; i.e., typically, a larger lot sells for a lower price per square meter than for a smaller lot. In this section, we will attempt to determine whether this is true for land plots in Tokyo by allowing the cost of land to be a *piecewise linear function* of the area of the land that the structure sits on.[24] Another possible limitation of our model is that the assumption of a straight line (net) depreciation rate for all ages of a residential dwelling may not be true. Thus in this section, we will attempt to increase the descriptive power of Model 1 by allowing the net depreciation of the structure to be a *piecewise linear function* of the age of the structure.[25]

We first consider how to model possible nonlinearities in the price of residential land. We divide up our 5578 observations into 3 roughly equal groups of observations based on their lot sizes. Recall that we have restricted the range of the land variable to $0.5 \le L_{tn} \le 2.5$.[26] We chose the land areas where there is a change in the marginal price of land to be $L_1 \equiv 0.77$ and $L_2 \equiv 1.10$. Using these *land break points*, we found that 1861 observations fell into the interval $0.5 \le L_{tn} < 0.77$, 1833 observations fell into the interval $0.77 \le L_{tn} < 1.10$ and 1884 observations fell into the interval $1.1 \le L_{tn} \le 2.5$.[27] We label the three sets

---

[23] The mean and median series cannot adjust properly for changes in the relative prices of land and structures or for changes in the average age of the houses sold. Also our mean and median series are for all sales of houses in Tokyo and thus these series were not adjusted for changes in the number of properties sold in expensive wards and less expensive wards. We cannot expect the mean and median series to be very accurate constant quality indexes of house prices; see de Haan and Diewert (2011).

[24] For the town of "A" in the Netherlands, Diewert, de Haan and Hendriks (2011a) (2011b) found that the marginal price of land rose for medium size lots and then fell for very large lots. These papers used the linear spline model for lot size that we will use in this section.

[25] In the statistics literature, models that make the dependent variable in a regression model a piecewise linear function of an exogenous variable are called *linear spline models*. Diewert (2003a; 328-329) proposed the type of nonlinear hedonic regression model defined by (9) and discussed its flexibility properties.

[26] Recall that our units of measurement for land are in 100 meters squared so that $L_{tn} = 1$ means that observation n in period t had a land area equal to 100 m^2.

[27] Thus the sample probabilities for an observation to fall into the 3 land intervals are 0.33363, 0.32861 and 0.33776.

of observations that fall into the above three groups as groups 1-3. For each observation n in period t, we define the three *land dummy variables*, $D_{L,tn,k}$, for k = 1,2,3 as follows:[28]

(9) $D_{L,tn,k} \equiv 1$ 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.

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

$$(10)\ f_L(L_{tn}) \equiv D_{L,tn,1}\lambda_1 L_{tn} + D_{L,tn,2}[\lambda_1 L_1 + \lambda_2(L_{tn}-L_1)] + D_{L,tn,3}[\lambda_1 L_1 + \lambda_2(L_2-L_1) + \lambda_3(L_{tn}-L_2)]$$

where the λ_k are unknown parameters and $L_1 \equiv 0.77$ and $L_2 \equiv 1.10$. The function $f_L(L_{tn})$ defines a *relative valuation function for the land area of a house* as a function of the plot area. Thus if $0.5 \le L_{tn} < 0.77$, then the relative land value of observation n in period t is $f_L(L_{tn}) = \lambda_1 L_{tn}$; if $0.77 \le L_{tn} < 1.10$, then the relative land value of observation n in period t is $f_L(L_{tn}) = \lambda_1 L_1 + \lambda_2(L_{tn}-L_1)$ and if $1.1 \le L_{tn} \le 2.5$, then the relative land value of observation n in period t is $f_L(L_{tn}) = \lambda_1 L_1 + \lambda_2(L_2-L_1) + \lambda_3(L_{tn}-L_2)$. If observation n in period t is in Ward 10, then we will set the land value of this house equal to $\alpha_t f_L(L_{tn})$.

We turn our attention to modeling possible nonlinearities in the net depreciation rate. We again attempt to divide up our 5578 observations into 3 roughly equal groups based on the age of the structure. Recall that we have restricted the range of the age variable to $0 \le A_{tn} \le 50$. We chose the house ages where there is a change in the marginal depreciation rate to be $A_1 \equiv 10$ and $A_2 \equiv 20$. Using these *age break points*, we found that 2085 observations fell into the interval $0 \le A_{tn} < 10$, 1996 observations fell into the interval $10 \le A_{tn} < 20$ and 1497 observations fell into the interval $20 \le A_{tn} \le 50$.[29] We label the three sets of observations that fall into the above three groups as groups 1-3. For each observation n in period t, we define the three *Age dummy variables*, $D_{A,tn,m}$, for m = 1,2,3 as follows:[30]

(11) $D_{A,tn,m} \equiv 1$ if observation tn has a structure whose age belongs to group m;
$\equiv 0$ if observation tn has a structure whose age is *not* in group m.

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

$$(12)\ g_A(A_{tn}) \equiv 1 - \{D_{A,tn,1}\delta_1 A_{tn} + D_{A,tn,2}[\delta_1 A_1 + \delta_2(A_{tn}-A_1)] + D_{A,tn,2}[\delta_1 A_1 + \delta_2(A_2-A_1) + \delta_3(A_{tn}-A_2)]\}$$

---

[28] Note that for each observation, the land dummy variables sum to one; i.e., for each tn, $D_{L,tn,1} + D_{L,tn,2} + D_{L,tn,3} = 1$.
[29] Thus the sample probabilities for an observation to fall into the 3 age intervals are 0.37379, 0.35783 and 0.26838.
[30] Note that for each observation, the Age dummy variables sum to one; i.e., for each tn, $D_{A,tn,1} + D_{A,tn,2} + D_{A,tn,3} = 1$.

where the δ_k are unknown parameters and $A_1 \equiv 10$ and $A_2 \equiv 20$. The function $g_a(A_{tn})$ defines a (relative) *depreciation schedule for a house structure* as a function of the structure age. Consider house n that sold in period t. If the age of the structure is 0 years so that it is a new structure, then its relative value is set equal to 1. If $0 < A_{tn} < 10$, then its structure value relative to a brand new structure is set equal to $g_A(A_{tn}) \equiv 1 - \delta_1 A_{tn}$. If $10 \leq A_{tn} < 20$, then its relative structure value is set equal to $g_A(A_{tn}) \equiv 1 - \delta_1 A_1 - \delta_2(A_{tn} - A_1)$. Finally, if $20 \leq A_{tn} \leq 50$, then its relative structure value is set equal to $g_A(A_{tn}) \equiv 1 - \delta_1 A_1 - \delta_2(A_2 - A_1) - \delta_3(A_{tn} - A_2)$. Thus the depreciation schedule for a house is now a piecewise linear schedule as opposed to the linear or straight line schedule that was used in the previous section.[31]

Now we are ready to define our new nonlinear regression model that generalizes the model defined by (5) and (6). For t = 1,...,44 and n = 1,...,N(t):

$$(13)\ V_{tn} = \alpha_t\{\textstyle\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn}) + \beta p_{Ct}\, g_A(A_{tn}) S_{tn} + \varepsilon_{tn}$$

where the functions f_L and g_A are defined above by (10) and (12) and ε_{tn} is an error term. There are 44 unknown land price parameters α_t, 1 structure price level parameter β, 21 ward relative land price level parameters ω_j, 3 lot size parameters λ_k and three depreciation parameters δ_m to estimate. However, as was the case with Model 1, not all parameters in (11) can be identified. Hence we impose the following identifying restrictions on the parameters:

$$(14)\ \omega_{10} = 1;\ \lambda_1 = 1.$$

Thus there are 44+1+20+2+3 = 70 unknown parameters to be estimated. The nonlinear regression model defined by (11) and (12) is our *Model 2*.

As was the case with Model 1, we estimated the parameters for Model 2 using the nonlinear regression option in Shazam.[32] The detailed parameter estimates are listed in the Appendix in Table A3.[33] The R^2 for this model turned out to be 0.8206 and the log likelihood was −9164.1, an increase of 68.9 over the Model 1 log likelihood.[34] Thus

---

[31] Note that if $\delta_1 = \delta_2 = \delta_3$, then the present depreciation model reduces to straight line depreciation. If in addition, $\lambda_1 = \lambda_2 = \lambda_3$, then the nonlinear regression model in this section reduces to the model in the previous section.

[32] Each of the four models that we propose in this paper subsequent to the first model is a generalization of the previous model so we were able to use the final estimates of the previous model as starting values for the parameters of each new model to facilitate convergence of the nonlinear estimation. No convergence difficulties were encountered.

[33] We note that the annual net depreciation rate for Model 1 was estimated as $\delta = 1.39\%$ with a T statistic of 26.8.

[34] The sum of the residuals in this model was only −0.5, a negligible amount. Thus adding a constant term to the regression would not add significantly to the fit of Model 2. We did not include a constant term in the regression because we want to allocate the value of the sale to separate land and structure components that add up to the total sale value. We note that the residual sum in Model 1 was 165.5 so Model 2 is much better in this respect.

adding the 2 extra lot size parameters and the 2 extra depreciation parameters is well justified.

Recall that we set λ_1 equal to 1 and the estimated λ_2 and λ_3 turned out to be 0.7533 and 0.9486 respectively. The interpretation of these parameters runs as follows. If observation n in period t had a land area L_{tn} which was less than L_1 = .77 (which is 77 m^2 since we are measuring land area in units of 100 m^2) and it was located in Ward 10, then its estimated land value is $\alpha_t\lambda_1 L_{tn} = \alpha_t L_{tn}$ However, if the land area was between L_1 and L_2 = 1.1 (110 m^2), then its estimated land value is $\alpha_t[L_1+\lambda_2(L_{tn}-L_1)]$. Thus the relative (to α_t) marginal price of land shifts from λ_1 =1 until L_{tn} reaches the land level L_1, and then for amounts of land beyond this level (but less than the level defined by L_2), the relative marginal price of land is λ_2 = 0.7533 according to our estimated coefficient. If the land area of observation tn was greater than or equal to L_2, then its estimated land value is $\alpha_t[L_1+\lambda_2(L_2-L_1)+\lambda_3(L_{nt}-L_2)]$. Thus the relative marginal price of land shifts from λ_2 to λ_3 for plot areas greater than or equal to L_3 = 1.1 (110 m^2). Our estimate for the relative marginal price of land for large lots is λ_3 = 0.9486. Note that these same relative marginal valuations for land apply to all periods t; i.e., the period t land price parameter α_t shifts the entire schedule of land values as a function of land size in a proportional manner for each period t. Thus normalizing on the price of land for small lots, we find that for lots of medium size, the relative marginal price of land falls from 1 to 0.7533 for land areas between L_1 and L_2 and for larger lots greater than L_2, the relative marginal price of land increases to 0.9486. Thus in any given period, the estimated value of the land component of the housing sale is a continuous piecewise linear function of the lot size.

The estimated value of (net) depreciation also follows a piecewise linear schedule instead of just being a linear function of age as in Model 1. Our estimated net depreciation rate parameters for Model 2 were δ_1 = 0.0247, δ_2 = 0.0159 and δ_3 = 0.0032. To explain the meaning of these parameters, consider an observation n in period t that has house age equal to A_{tn} years. If $0 \le A_{nt} < A_1 \equiv 10$ years, then our estimated net depreciation of the house in terms of the period t price of a unit of new house construction, βp_{Ct}, is $\beta p_{Ct}\delta_1 A_{tn}$. Thus for relatively new houses, we have a simple straight line depreciation model (in terms of current structure prices) and the annual net depreciation rate for these relatively new houses is 2.47% per year. However, if $A_1 \equiv 10 \le A_{tn} < 20 \equiv A_2$ so that the age of the house is between 10 and 20 years old, then our estimate for the net depreciation of the house in current period prices is $\beta p_{Ct}[\delta_1 A_1+\delta_2(A_{tn}-A_1)]$. Thus for this age group of houses sold, the marginal rate of net depreciation falls to 1.59% per year for ages A_{tn} greater than 10 years. Finally, if the age of the house is between $A_2 \equiv 20$ and 50 years old, then our estimate for the net depreciation of the house in current period prices is $\beta p_{Ct}[\delta_1 A_1+\delta_2(A_2-A_1)+\delta_3(A_{tn}-A_2)]$. Thus for this age group of houses sold, the marginal rate of net depreciation falls to 0.32% per year for A_{tn} greater than 20 years.[35]

[35] We conjecture that the reason why the marginal net depreciation rate for houses older than 20 years is so low is that houses that survive beyond 20 years of age have been extensively renovated or are heritage houses. We are estimating net depreciation rates here because we have no information on the magnitude of renovation expenditures.

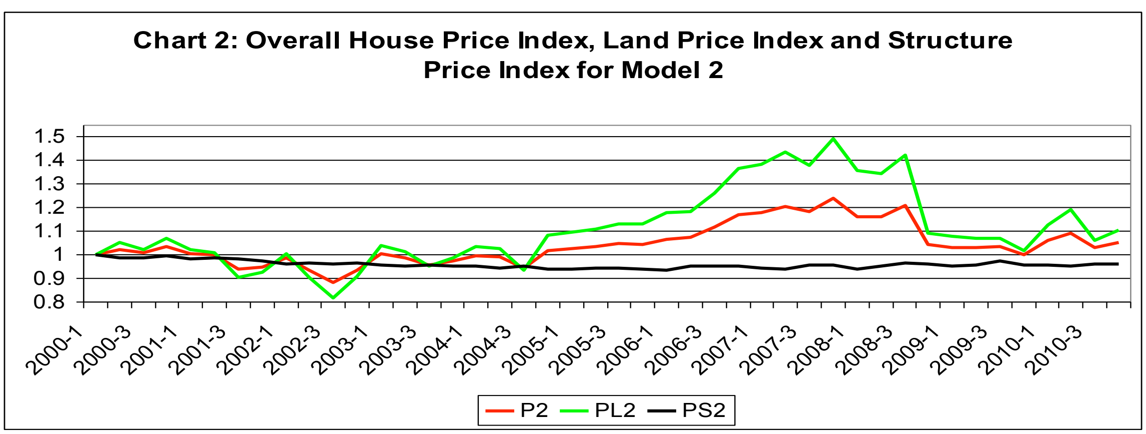
Model 2 defined by (13) and (14) decomposes into two terms: one which involves the land area L_{tn} of the house and another which involves the structure area S_{tn} of the house. As was the case with Model 1, the first term can be regarded as an estimate of the land value of house n that was sold in quarter t while the second term is an estimate of the structure value of the house. We follow the same strategy in decomposing the land and structure values into price and quantity components as in Model 1. The quarterly time coefficients α_t act as proportional time shifters of the hedonic surface for the land component of each house in our sample and the relative period t to period s land price for each house is α_t/α_s. As was the case with Model 1, the quarterly time coefficients βp_{Ct} act as proportional time shifters of the hedonic surface for the structure component of each house in our sample and the period t to period s land price for each house in our sample again turns out to be p_{Ct}/p_{Cs}.

Thus the Model 2 *constant quality residential land price index* for Tokyo for quarter t is defined to be $P_{L2t} \equiv \alpha_t/\alpha_1$ and the corresponding *constant quality residential structures price index* for Tokyo for quarter t is defined to be $P_{S2t} \equiv p_{Ct}/p_{C1}$.[36] The corresponding Model 2 quarter t *constant quality quantity levels*, Q_{L2t} and Q_{S2t}, are defined as the total quarter t values of land and structures divided by the corresponding price levels for t = 1,...,44:

(15) $Q_{L2t} \equiv \sum_{n=1}^{N(t)} \alpha_1\{\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn})$;
(16) $Q_{S2t} \equiv \sum_{n=1}^{N(t)} \beta p_{Ct}\, g_A(A_{tn}) S_{tn}$.

We again use the Fisher ideal index to aggregate the price and quantity components for land and structures into a house price index. Thus define the *overall house price level for quarter t* for Model 2, P_{2t}, as the chained Fisher price index of the land and structure series $\{P_{L2t},P_{S2t},Q_{L2t},Q_{S2t}\}$.

The overall Model 2 house price index P_{2t} as well as the land and structure price indexes P_{L2t} and P_{S2t} for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 2 below.[37]



[36] Note that $P_{S1t} = P_{S2t}$.
[37] The series P_2, P_{L2} and P_{S2} are also listed in Table A4 of the Appendix.

From Chart 2, it can be seen that there was a mini land price bubble during the years 2006-2008 for residential properties in Tokyo. Comparing Charts 1 and 2, it can be seen that the structure price index is the same in both Models (by construction) and the land and overall indexes are much the same in both Models.[38]

In the following section, we will generalize Model 2 by adding some additional explanatory variables that are thought to be important in explaining house price movements in Tokyo.

5. Quality Adjustment for the Number of Bedrooms and Lot Width

Many hedonic regression models that attempt to explain movements in house prices use the number of rooms or bedrooms in the structure as an explanatory variable. We will use the *number of bedrooms*, NB_{tn}, for house n sold in period t as a quality adjusting variable for the structure. In Japan, the *width of the lot*, WI_{tn}, is also thought to be an important characteristic that explains the value of a residential property (a bigger width is thought to more desirable).

We will treat the number of bedrooms variable in a manner that is similar to our treatment of depreciation. We first need to break up our sample into three groups of observations: houses with a low number of bedrooms, houses with a medium number and houses with a high number of bedrooms. We find that there are 247 houses with 2 bedrooms, 1628 with 3 bedrooms, 2439 with 4 bedrooms and 1264 houses with 5-8 bedrooms. We will allocate the 2 and 3 bedroom houses to the low group, the 4 bedroom houses to the medium group and the 5-8 bedroom houses to the high group. We transform the number of bedrooms variable, NB, into the number of bedrooms less 2 variable B; i.e., for observation n in period t, define the *translated number of bedrooms variable* B_{tn} as follows:

(17) $B_{tn} \equiv NB_{tn} - 2$; $\quad t = 1,...,44$; $n = 1,...,N(t)$.

Thus the B variable takes on integer values between 0 and 6. If B_{tn} equals 0 or 1, then observation tn falls into the low number of bedrooms group. If $B_{tn} = 2$, then observation tn falls into the medium number of bedrooms group. If $B_{tn} = 3$-6, then observation tn falls into the high number of bedrooms group. The *break points* for the B variable where there is a change in the marginal value of extra bedrooms are chosen to be $B_1 \equiv 1$ and $B_2 \equiv 2$. The *bedroom dummy variables*, $D_{B,tn,k}$, are defined as follows:

(18) $D_{B,tn,1} \equiv 1$ if $B_{tn} = 0$ or 1; $D_{B,tn,1} \equiv 0$ if $B_{tn} > 1$;
$D_{B,tn,2} \equiv 1$ if $B_{tn} = 2$; $\quad D_{B,tn,1} \equiv 0$ if $B_{tn} \neq 2$;
$D_{B,tn,1} \equiv 1$ if $B_{tn} > 2$; $\quad D_{B,tn,1} \equiv 0$ if $B_{tn} \leq 1$.

Now consider the following piecewise linear function of B_{tn}, $g_B(B_{tn})$, defined as follows:

[38] The correlation coefficients between P_1 and P_2 and P_{L1} and P_{L2} were 0.99941 and 0.99946 respectively.

$$(19)\ g_B(B_{tn}) \equiv \phi_1 + D_{B,tn,1}\phi_2 B_{tn} + D_{B,tn,2}[\phi_2 B_1 + \phi_3(B_{tn} - B_1)] + D_{B,tn,3}[\phi_2 B_1 + \phi_3(B_2 - B_1) + \phi_4(B_{tn} - B_2)]$$

where the ϕ_k are unknown parameters and $B_1 \equiv 1$ and $B_2 \equiv 2$. Thus if $B_{tn} = 0$ (so that house n sold in period t has 2 bedrooms), then $g_B(B_{tn}) = g_B(0) = \phi_1$. If $B_{tn} = 1$ (so that house n sold in period t has 3 bedrooms), then $g_B(B_{tn}) = g_B(1) = \phi_1 + \phi_2$. If $B_{tn} = 2$ (so that house n sold in period t has 4 bedrooms), then $g_B(B_{tn}) = g_B(2) = \phi_1 + \phi_2 + \phi_3$. Finally, if B_{tn} = 3-6 (so that house n sold in period t has 5-8 bedrooms), then $g_B(B_{tn}) = \phi_1 + \phi_2 + \phi_3 + \phi_4(B_{tn} - 2)$. We will use the function g_B to determine the *relative value of a house as a function of the number of bedrooms* that it has, holding other characteristics constant. It can be seen that this function is a linear spline function and is relatively flexible in that it can describe a large number of structure valuations with different choices of the 4 ϕ_k parameters.[39]

We turn now to our parameterization of the relative value of the land area of a house as a function of the lot width WI (or frontage). Recall that the width variable ranged between 2.5 and 9 meters. We transform the width variable to the width variable less 2.5; ; i.e., for observation n in period t, define the *translated frontage variable* F_{tn} as follows:

$$(20)\ F_{tn} \equiv WI_{tn} - 2.5\ ; \qquad t = 1,...,44\ ;\ n = 1,...,N(t).$$

Thus the range of F_{tn} is $0 \leq F_{tn} \leq 6.5$. We will use a relative valuation model for lots of different widths similar to the above relative valuation model for the number of bedrooms. We chose the frontage widths where there is a change in the marginal valuation of translated width to be $F_1 \equiv 1.5$ and $F_2 \equiv 2.5$. Using these *width break points*, we found that 1109 observations fell into the interval $0 \leq F_{tn} < 1.5$, 2352 observations fell into the interval $1.5 \leq F_{tn} < 2.5$ and 2117 observations fell into the interval $2.5 \leq F_{tn} \leq 6.5$.[40] We label the three sets of observations that fall into the above three groups as groups 1-3. For each observation n in period t, we define the three *frontage dummy variables*, $D_{F,tn,k}$, for k = 1,2,3 as follows:[41]

(21) $D_{F,tn,k} \equiv 1$ if observation tn has translated frontage width that belongs to group k;
$\equiv 0$ if observation tn has translated frontage width that does not belong to group k.

Now consider the following piecewise linear function of F_{tn}, $f_F(F_{tn})$, defined as follows:

$$(22)\ f_F(F_{tn}) \equiv \kappa_1 + D_{F,tn,1}\kappa_2 F_{tn} + D_{F,tn,2}[\kappa_2 F_1 + \kappa_3(F_{tn} - F_1)] + D_{F,tn,3}[\kappa_2 F_1 + \kappa_3(F_2 - F_1) + \kappa_4(F_{tn} - F_2)]$$

---

[39] We expect these parameters to be positive numbers.

[40] Thus the sample probabilities for an observation to fall into the 3 lot width intervals are 0.19882, 0.42166 and 0.37953.

[41] Note that for each observation, the frontage width dummy variables sum to one; i.e., for each tn, $D_{F,tn,1} + D_{F,tn,2} + D_{F,tn,3} = 1$.

where the κ_k are unknown parameters and $F_1 = 1.5$ and $F_2 = 2.5$. If $F_{tn} < 1.5$, then $f_F(F_{tn}) = \kappa_1+\kappa_2F_{tn}$. If $1.5 \leq F_{tn} < 2.5$, then $f_F(F_{tn}) = \kappa_1+\kappa_2F_1+\kappa_3(F_{tn}-F_1)$. If $2.5 \leq F_{tn}$, then $f_F(F_{tn}) = \kappa_1+\kappa_2F_1+\kappa_3(F_2-F_1)+\kappa_4(F_{tn}-F_2)$. We will use the piecewise linear function f_F to determine the *relative value of the land area as a function of the width of the lot*, holding other characteristics constant.

Noting that the number of bedrooms is a characteristic that may affect the value of the structure and the lot width is a characteristic that may affect the value of the land area that the structure sits on, we multiply the land value term for observation n in period t in Model 2 by $f_W(W_{tn})$ and the corresponding structure value by $g_B(B_{tn})$. This leads to the following nonlinear regression model for t = 1,...,44 and n = 1,...,N(t):

(23) $V_{tn} = \alpha_t\{\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn})f_F(F_{tn}) + \beta p_{Ct}\, g_A(A_{tn})g_B(B_{tn})S_{tn} + \varepsilon_{tn}$

where the functions f_L, g_A, g_B and f_F are defined above by (10), (12), (19) and (22) respectively. There are 44 unknown land price parameters α_t, 1 structure price level parameter β, 21 ward relative land price level parameters ω_j, 3 lot size parameters λ_k, three depreciation parameters δ_m, 4 number of bedroom parameters ϕ_k and 4 frontage width parameters κ_k to estimate. However, as was the case with Models 1 and 2, not all parameters in (23) can be identified. Hence we impose the following identifying restrictions on the parameters:

(24) $\omega_{10} = 1$; $\lambda_1 = 1$; $\phi_1 = 1$ and $\kappa_1 = 1$.

Thus there are 44+1+20+2+3+3+3 = 76 unknown parameters to be estimated. The nonlinear regression model defined by (23) and (24) is our *Model 3*.

We estimated the unknown parameters for Model 3 using the nonlinear regression option in Shazam.[42] The detailed parameter estimates are listed in the Appendix in Table A5. The R^2 for this model turned out to be 0.8256 and the log likelihood was −9085.3, an increase of 78.7 over the Model 2 log likelihood.[43] Thus adding the 3 extra lot width parameters and the 3 extra bedroom parameters is well justified.

The estimated lot width parameters were $\kappa_2 = 0.1038$, $\kappa_3 = 0.0433$ and $\kappa_4 = 0.0124$. The interpretation of these parameters runs as follows: for properties in the small lot frontage width group, an extra meter of lot width adds 10.38% to the land value; for properties in the medium lot with group, an extra meter of lot width adds 4.33% to the land value and properties in the large lot width group, an extra meter of lot width adds 1.24% to the land value of the property. Thus there are diminishing returns to lot width but extra lot width (holding other characteristics constant) always adds to the land value of the property.

[42] Each of the four models that we propose in this paper subsequent to the first model is a generalization of the previous model so we were able to use the final estimates of the previous model as starting values for the parameters of each new model to facilitate convergence of the nonlinear estimation. No convergence difficulties were encountered.

[43] The sum of the residuals in this model was −20.8.

The estimated number of bedroom parameters were $\phi_2 = 0.0277$, $\phi_3 = -0.0326$ and $\phi_4 = -0.0437$.[44] The economic interpretation of these parameters is as follows: holding all other characteristics constant (including the size of the structure in meters squared), going from 2 bedrooms to 3 bedrooms *adds* 2.77% to the value of the structure; going from 3 bedrooms to 4 bedrooms *subtracts* 3.26% from the value of the structure and for each bedroom beyond 4 bedrooms, *subtract* 4.37% from the value of the structure. These results are a bit hard to interpret because they are conditional on the area of the structure. Now for small structures, we would expect that the "optimal" number of bedrooms is small and for large structures, we would expect that the "optimal" number of bedrooms is large. However, our very simple model makes a bedroom value adjustment over all structure sizes and so our interpretation of the above numerical results is that for a structure of average size in terms of its floor space area, it is preferable to have 3 bedrooms over 2 but beyond 3 bedrooms, for an average sized house, adding more bedrooms subtracts from the value of the property.[45]

Model 3 defined by (23) and (24) decomposes into two terms: one which involves the land area L_{tn} of the house and another which involves the structure area S_{tn} of the house. As was the case with Models 1 and 2, the first term can be regarded as an estimate of the land value of house n that was sold in quarter t while the second term is an estimate of the structure value of the house. We follow the same strategy in decomposing the land and structure values into price and quantity components as in the previous Models. The quarterly time coefficients α_t act as proportional time shifters of the hedonic surface for the land component of each house in our sample and the relative period t to period s land price for each house is α_t/α_s. As was the case with Model 1, the quarterly time coefficients βp_{Ct} act as proportional time shifters of the hedonic surface for the structure component of each house in our sample and the period t to period s land price for each house in our sample again turns out to be p_{Ct}/p_{Cs}.

Thus the Model 3 *constant quality residential land price index* for Tokyo for quarter t is defined to be $P_{L3t} \equiv \alpha_t/\alpha_1$ and the corresponding *constant quality residential structures price index* for Tokyo for quarter t is defined to be $P_{S3t} \equiv p_{Ct}/p_{C1}$. The corresponding Model 3 quarter t *constant quality quantity levels*, Q_{L3t} and Q_{S3t}, are defined as the total quarter t values of land and structures divided by the corresponding price levels for t = 1,...,44:

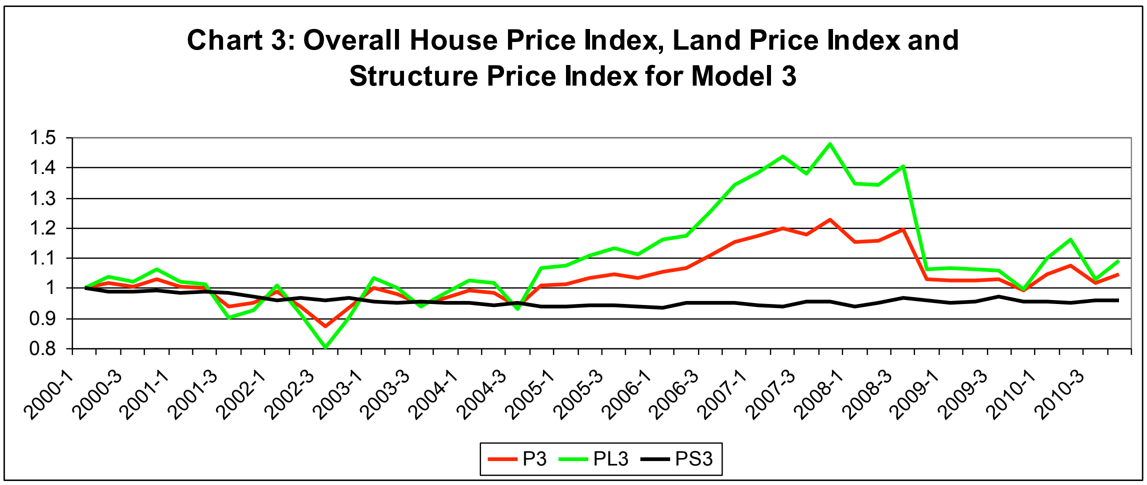
[44] The T statistics for these 3 parameters were 0.744, −2.227 and −6.051.

[45] Looking at the model defined by (23), it can be seen that we have assumed that the structure value of the property n in period t, $\beta p_{Ct} g_A(A_{tn}) g_B(B_{tn}) S_{tn}$, is basically the structure area S_{tn} times a period t structure price parameter βp_{Ct}, times some quality adjustment factors that depend on the age of the structure, $g_A(A_{tn})$, and the number of bedrooms in the structure, $g_B(B_{tn})$. Thus we are assuming that these quality adjustment factors act more or less independently of each other in a multiplicative fashion; i.e., we have a kind of multiplicative separability (or statistical independence) assumption. This type of model can provide a first order approximation to a more general hedonic surface in time, age, the number of bedrooms and the floor space area. However, in order to capture adequately the interaction effects of B_{tn} and S_{tn}, we would require a functional form that could provide a second order approximation. In this paper, we did not venture beyond hedonic surface functional forms that can provide a first order approximation. First and second order approximation properties of hedonic functional forms was discussed by Diewert (2003a; 329-334).

(25) $Q_{L3t} \equiv \sum_{n=1}^{N(t)} \alpha_1 \{\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn}) f_F(F_{tn})$;
(26) $Q_{S3t} \equiv \sum_{n=1}^{N(t)} \beta p_{Ct}\, g_A(A_{tn}) g_B(B_{tn}) S_{tn}$.

We again use the Fisher ideal index to aggregate the price and quantity components for land and structures into a house price index. Thus define the *overall house price level for quarter t* for Model 3, P_{3t}, as the chained Fisher price index of the land and structure series $\{P_{L3t},P_{S3t},Q_{L3t},Q_{S3t}\}$.

The overall Model 3 house price index P_{3t} as well as the land and structure price indexes P_{L3t} and P_{S3t} for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 3 below.[46]



Comparing Charts 1, 2 and 3, it can be seen that the structure price index is the same in both Models (by construction) and the land and overall indexes are much the same in all three Models.[47]

In the following section, we will generalize Model 3 by adding some additional explanatory variables that are important in explaining house price movements in Tokyo.

6. Quality Adjustment for the Nearness to Subway Lines and Subway Travel Time

Recall that in section 2, we noted that we constructed information on the variables TW and TT for each house in our sample. TW_{tn} is the time in minutes it takes to walk from house n sold in period t to the nearest subway station while TT_{tn} is the time in minutes the train takes from the nearest subway station to the main Tokyo station. Recall that the sample range of TW was 2 to 29 minutes while the sample range of TT was 4 to 48 minutes.

[46] The series P_3, P_{L3} and P_{S3} are also listed in Table A6 of the Appendix.
[47] The correlation coefficients between P_3 and P_1 and P_2 were 0.99678 and 0.99811 respectively and P_{L3} and P_{L1} and P_{L2} were 0.99689 and 0.99806 respectively.

We transform the TW variable to the TW variable less 2; ; i.e., for observation n in period t, define the *transformed walking time variable* M_{tn} as follows:

(27) $M_{tn} \equiv TW_{tn} - 2$; $\quad t = 1,...,44$; $n = 1,...,N(t)$.

Thus the range of M_{tn} is $0 \leq M_{tn} \leq 27$. As usual, we want to group the properties in our sample into 3 groups of roughly equal size. We chose our break points for M to be $M_1 \equiv 6$ and $M_2 \equiv 11$. Using these break points, we found that 1811 observations fell into the interval $0 \leq M_{tn} < 6$, 2261 observations fell into the interval $6 \leq M_{tn} < 11$ and 1506 observations fell into the interval $11 \leq M_{tn} \leq 27$.[48] We label the three sets of observations that fall into the above three groups as groups 1-3. For each observation n in period t, we define the three *time to nearest subway station dummy variables*, $D_{M,tn,k}$, for k = 1,2,3 as follows:[49]

(28) $D_{M,tn,k} \equiv 1$ if observation tn has translated subway walking time that belongs to group k;
$\equiv 0$ if observation tn has translated subway walking time that does not belong to group k.

Now consider the following piecewise linear function of M_{tn}, $f_M(M_{tn})$, defined as follows:

(29) $f_M(M_{tn}) \equiv \tau_1 + D_{M,tn,1}\, \tau_2 M_{tn} + D_{M,tn,2}[\tau_2 M_1 + \tau_3(M_{tn} - M_1)]$
$+ D_{M,tn,3}[\tau_2 M_1 + \tau_3(M_2 - M_1) + \tau_4(M_{tn} - M_2)]$

where the τ_k are unknown parameters and $M_1 = 6$ and $M_2 = 11$. If $M_{tn} < 6$, then $f_M(M_{tn}) = \tau_1 + \tau_2 M_{tn}$. If $6 \leq W_{tn} < 11$, then $f_M(M_{tn}) = \tau_1 + \tau_2 M_1 + \tau_3(M_{tn} - M_1)$. If $11 \leq M_{tn}$, then $f_M(M_{tn}) = \tau_1 + \tau_2 M_1 + \tau_3(M_2 - M_1) + \tau_4(M_{tn} - M_2)$. We will use the piecewise linear function f_M to determine the *relative value of the land area as a function of the travel time to the nearest subway station*, holding other characteristics constant.

The subway travel time variable TT has the range 4 to 48 minutes. We translate this variable to start at zero. Thus define the *translated subway travel time variable* T_{tn} as follows:

(30) $T_{tn} \equiv TT_{tn} - 4$; $\quad t = 1,...,44$; $n = 1,...,N(t)$.

The range of T_{tn} is $0 \leq T_{tn} \leq 44$. As usual, we want to group the properties in our sample into 3 groups of roughly equal size. We chose our break points for T to be $T_1 \equiv 24$ and $T_2 \equiv 32$. Using these break points, we found that 1678 observations fell into the interval $0 \leq T_{tn} < 24$, 2049 observations fell into the interval $24 \leq T_{tn} < 32$ and 1851 observations fell

[48] Thus the sample probabilities for an observation to fall into the 3 (translated) time to nearest subway station groups are 0.32467, 0.40534 and 0.26999.

[49] Note that for each observation, the subway time dummy variables sum to one; i.e., for each tn, $D_{M,tn,1} + D_{M,tn,2} + D_{M,tn,3} = 1$.

into the interval $32 \leq T_{tn} \leq 44$.[50] We label the three sets of observations that fall into the above three groups as groups 1-3. For each observation n in period t, we define the three *travel time to Tokyo station dummy variables*, $D_{T,tn,k}$, for k = 1,2,3 as follows:

(31) $D_{T,tn,k} \equiv 1$ if observation tn has translated time to Tokyo station that belongs to group k;
$\equiv 0$ if observation tn has translated time to Tokyo station that does not belong to group k.

Now consider the following piecewise linear function of T_{tn}, $f_T(T_{tn})$, defined as follows:

(32) $f_T(T_{tn}) \equiv \mu_1 + D_{T,tn,1}\,\mu_2 T_{tn} + D_{T,tn,2}[\mu_2 T_1 + \mu_3(T_{tn} - T_1)]$
$+ D_{T,tn,3}[\mu_2 T_1 + \mu_3(T_2 - T_1) + \mu_4(T_{tn} - T_2)]$

where the μ_k are unknown parameters and $T_1 = 24$ and $T_2 = 32$. If $T_{tn} < 24$, then $f_T(T_{tn}) = \mu_1 + \mu_2 T_{tn}$. If $24 \leq T_{tn} < 32$, then $f_T(T_{tn}) = \mu_1 + \mu_2 T_1 + \mu_3(T_{tn} - T_1)$. If $32 \leq T_{tn}$, then $f_T(T_{tn}) = \mu_1 + \mu_2 T_1 + \mu_3(T_2 - T_1) + \mu_4(T_{tn} - T_2)$. We will use the piecewise linear function f_T to determine the *relative value of the land area as a function of the travel time from the nearest subway station to the Tokyo Central station*, holding other characteristics constant.

The travel time characteristics are ones that may affect the value of the land that a house sits on. Thus we multiply the land value term in Model 3 for observation n in period t by $f_M(M_{tn})f_T(T_{tn})$. This leads to the following nonlinear regression model for t = 1,...,44 and n = 1,...,N(t):

(33) $V_{tn} = \alpha_t\{\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn}) + \beta p_{Ct}\, g_A(A_{tn}) g_B(B_{tn}) S_{tn} + \varepsilon_{tn}$

where the functions f_L, g_A, g_B, f_F, f_M and f_T are defined above by (10), (12), (19), (22), (29) and (32) respectively. Compared to the previous Model, we have added 8 new subway time parameters, the 4 walking time parameters τ_k and the 4 subway travel time to the Tokyo station parameters μ_k, for a total of 88 parameters. However, as was the case with our previous models, not all parameters in (33) can be identified. Hence we impose the following identifying restrictions on the parameters:

(34) $\omega_{10} = 1$; $\lambda_1 = 1$; $\phi_1 = 1$; $\kappa_1 = 1$; $\tau_1 = 1$ and $\mu_1 = 1$.

Thus there are 82 unknown parameters to be estimated. The nonlinear regression model defined by (33) and (34) is our *Model 4*.

As usual, we estimated the unknown parameters for Model 4 using the nonlinear regression option in Shazam. The detailed parameter estimates are listed in the Appendix in Table A7. The R^2 for this model turned out to be 0.8417 and the log likelihood was

[50] Thus the sample probabilities for an observation to fall into the 3 (translated) travel time to the Tokyo station time groups are 0.30082, 0.36734 and 0.33184.

−8815.9, a very large increase of 269.4 over the Model 3 log likelihood.[51] Thus adding the 3 extra walking time parameters and the 3 extra travel time to Tokyo station parameters provides a significant addition to the explanatory power of our hedonic regression model.

The estimated walking time to the nearest subway station parameters were $\tau_2 = -0.0035$, $\tau_3 = -0.0201$ and $\tau_4 = -0.0171$. The interpretation of these parameters runs as follows: for properties where the walk to the nearest subway station is 2-8 minutes, an increase in walking time of 1 minute decreases the land value of the property by 0.35%; for properties where the walk to the nearest subway station is 8-13 minutes, an increase in walking time of 1 minute decreases the land value of the property by 2.01% and for properties where the walk to the nearest subway station is over 13 minutes, an increase in walking time of 1 minute decreases the land value of the property by 1.71%. Thus for properties that are quite close to a subway station, the drop in land value as walking time increases is not too substantial but as the walking time increases markedly, the drop in land value is quite substantial.

The estimated time from the nearest subway station to the Tokyo station parameters were $\mu_2 = -0.0008$, $\mu_3 = -0.0128$ and $\mu_4 = -0.0188$. The interpretation of these parameters runs as follows: for properties where the subway running time from the nearest subway station to the Tokyo station is 4-28 minutes, an increase in running time of 1 minute decreases the land value of the property by 0.08%, a negligible decease; for properties where the subway running time from the nearest subway station to the Tokyo station is 28-36 minutes, an increase in running time of 1 minute decreases the land value of the property by 1.28% and for properties where the subway running time from the nearest subway station to the Tokyo station is over 36 minutes, an increase in running time of 1 minute decreases the land value of the property by 1.88%, which is a substantial drop in value.

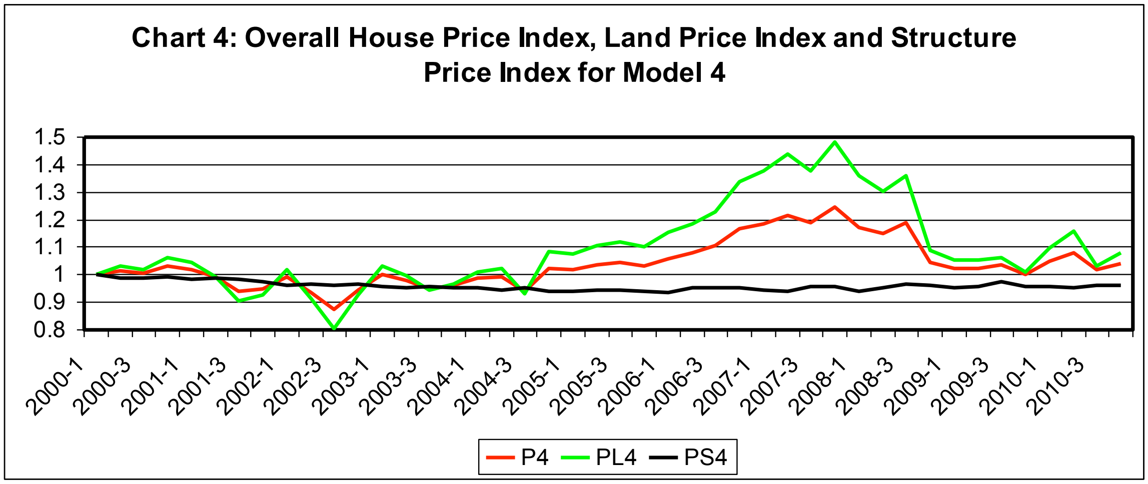
The Model 4 *constant quality residential land price index* for Tokyo for quarter t is defined to be $P_{L4t} \equiv \alpha_t/\alpha_1$ and the corresponding *constant quality residential structures price index* for Tokyo for quarter t is defined to be $P_{S4t} \equiv p_{Ct}/p_{C1}$. The corresponding Model 4 quarter t *constant quality quantity levels*, Q_{L4t} and Q_{S4t}, are defined as the total quarter t values of land and structures divided by the corresponding price levels for t = 1,...,44:

$$(35)\quad Q_{L4t} \equiv \sum_{n=1}^{N(t)} \alpha_1\{\sum_{j=1}^{21} \omega_j D_{W,tn,j}\} f_L(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn}) ;$$
$$(36)\quad Q_{S4t} \equiv \sum_{n=1}^{N(t)} \beta p_{Ct}\, g_A(A_{tn}) g_B(B_{tn}) S_{tn}.$$

We again use the Fisher ideal index to aggregate the price and quantity components for land and structures into a house price index. Thus define the *overall house price level for quarter t* for Model 4, P_{4t}, as the chained Fisher price index of the land and structure series $\{P_{L4t},P_{S4t},Q_{L4t},Q_{S4t}\}$.

[51] The sum of the residuals in this model was −11.4.

The overall Model 4 house price index P_{4t} as well as the land and structure price indexes P_{L4t} and P_{S4t} for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 4 below.[52]



Comparing Chart4 with the previous Charts, it can be seen that the structure price index is the same in all Models (by construction) and the land and overall indexes are much the same in all four Models.[53]

We have allowed for a different level of land prices across the 21 Wards in Tokyo that span our data set. However, we have forced all land prices to change proportionally across time with the estimated α_t being the proportional factors. It is unlikely that land prices in the different Wards move in an exactly fixed proportion. Thus in the following section, we divide up the Wards into two groups: those that have relatively high price levels (i.e., large estimated ω_j parameters) and those that have relatively low price levels (i.e., small ω_j parameters) and we allow land prices to move independently in these high and low end wards. We also allow the *level* of structure prices to differ in high and low end wards.

7. Allowing for Land and Structure Price Differences Across Wards

In many countries, property price movements differ substantially across expensive and less expensive neighborhoods. Usually, land price movements in high end properties are more volatile than in lower end properties. In this section, we will attempt to determine whether this pattern also holds for Tokyo residential land prices.

Ideally, it would be preferable to have separate land price parameters (the α_t) for each Ward. However, we do not have enough degrees of freedom to accurately measure land

---

[52] The series P_4, P_{L4} and P_{S4} are also listed in Table A8 of the Appendix.

[53] The correlation coefficients between P_4 and P_1, P_2 and P_3 were 0.99408, 0.99543 and 0.99643 respectively and the correlation coefficients between P_{L4} and P_{L1}, P_{L2} and P_{L3} were 0.99416, 0.99536 and 0.99650 respectively.

price movements ward by ward.[54] We do have a sufficient number of observations so that we can divide Wards into two groups based on the estimated ω_j parameters from Model 4: Group 1 Wards are those whose estimated relative land price levels ω_j exceeded 0.75 and Group 2 Wards are those whose estimated land price levels ω_j were less than 0.75. The following Wards were in Group 1 (the *expensive or high end Wards*): 1-4, 7-11, 13-14. The following Wards were in Group 2 (the *cheaper or lower end Wards*): 5, 12, 15-21. We will allow land prices to evolve over time in a completely independent manner for high and lower end Wards. Thus instead of estimating a single set of 44 land price parameters α_t, we will now estimate *two sets of land price parameters*: $\alpha_{1,t}$ for high end Wards and $\alpha_{2,t}$ for lower end Wards for t = 1,...,44.

Recall definition (10) which defined the quality adjustment for lot size function, $f_L(L_{tn})$. We will now allow for separate lot size quality adjustments in the high and lower end wards. The high and low end *lot size quality adjustment functions* for property n sold in period t, $f_{1L}(L_{tn})$ and $f_{2L}(L_{tn})$ respectively, are defined as follows for i = 1,2:

$$(37)\ f_{iL}(L_{tn}) \equiv D_{L,tn,1}\lambda_{i,1}L_{tn} + D_{L,tn,2}[\lambda_{i,1}L_1+\lambda_{i,2}(L_{tn}-L_1)] + D_{L,tn,3}[\lambda_{i,1}L_1+\lambda_{i,2}(L_2-L_1)+\lambda_{i,3}(L_{tn}-L_2)]$$

where the $\lambda_{i,k}$ are 6 unknown parameters, $L_1 \equiv 0.77$ and $L_2 \equiv 1.10$, and the lot size dummy variables $D_{L,tn,k}$ are defined above by (9). The parameters $\lambda_{1,1}$, $\lambda_{1,2}$ and $\lambda_{1,3}$ are the relative marginal prices of land for plots in high end wards and the parameters $\lambda_{2,1}$, $\lambda_{2,2}$ and $\lambda_{2,3}$ are the relative marginal prices of land for plots in lower end wards.

A final generalization over Model 4 is that we will now allow the level of structure prices to differ in high and lower end wards so that the previous structure price level parameter β is now replaced by β_1 (the level of structure prices in high end wards) and β_2 (the level of structure prices in lower end wards). Our expectation is that β_2 will be less than β_1 since we would expect the quality of construction to be higher in the high end wards. Our final nonlinear regression model is defined for t = 1,...,44 and n = 1,...,N(t) by the following equations:

$$\begin{aligned}(38)\ V_{tn} = {}& \alpha_{1,t}\{\omega_1 D_{W,tn,1}+\omega_2 D_{W,tn,2}+\omega_3 D_{W,tn,3}+\omega_4 D_{W,tn,4}+\omega_7 D_{W,tn,7}+\omega_8 D_{W,tn,8}+\omega_9 D_{W,tn,9} \\ & +\omega_{10} D_{W,tn,10}+\omega_{11} D_{W,tn,11}+\omega_{13} D_{W,tn,13}+\omega_{14} D_{W,tn,14}\} f_{1L}(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn}) \\ & +\alpha_{2,t}\{\omega_5 D_{W,tn,5}+\omega_6 D_{W,tn,6}+\omega_{12} D_{W,tn,12}+\omega_{15} D_{W,tn,15}+\omega_{16} D_{W,tn,16}+\omega_{17} D_{W,tn,17} \\ & +\omega_{18} D_{W,tn,18}+\omega_{19} D_{W,tn,19}+\omega_{20} D_{W,tn,11}+\omega_{21} D_{W,tn,21}\} f_{2L}(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn}) \\ & +\beta_1\{D_{W,tn,1}+D_{W,tn,2}+D_{W,tn,3}+D_{W,tn,4}+D_{W,tn,7}+D_{W,tn,8}+D_{W,tn,9}+D_{W,tn,10}+D_{W,tn,11} \\ & +D_{W,tn,13}+D_{W,tn,14}\} p_{Ct} g_A(A_{tn}) g_B(B_{tn}) S_{tn} \\ & +\beta_2\{D_{W,tn,5}+D_{W,tn,6}+D_{W,tn,12}+D_{W,tn,15}+D_{W,tn,16}+D_{W,tn,17}+D_{W,tn,18}+D_{W,tn,19} \\ & +D_{W,tn,20}+D_{W,tn,21}\} p_{Ct} g_A(A_{tn}) g_B(B_{tn}) S_{tn} + \varepsilon_{tn}\end{aligned}$$

[54] The total number of observations in Wards 1-21 were as follows: 69, 136, 82, 15, 32, 38, 144, 349, 409, 1158, 107, 305, 773, 124, 53, 34, 214, 925, 271, 143 and 197. Thus Wards 4, 5, 6 and 16 have only 15, 32, 33 and 34 observations. Thus the Wards with the most observations were Wards 10, 13 and 18 with 1158, 773 and 925 observations.

The explanatory variables on the right hand side of equations (38) decompose into 4 sets of terms:[55]

- The terms associated with $\alpha_{1,t}$ represent the estimated land value of a property in a high end ward;
- The terms associated with $\alpha_{2,t}$ represent the estimated land value of a property in a lower end ward;
- The terms associated with β_1 represent the estimated structure value of a property in a high end ward and
- The terms associated with β_2 represent the estimated structure value of a property in a lower end ward.

Compared to the previous Model, we have added 44 new land price parameters, $\alpha_{2,t}$, 3 new lot size quality adjustment parameters, $\lambda_{2,1}$, $\lambda_{2,2}$ and $\lambda_{2,3}$ and one new structure price level parameter, β_2. However, as was the case with our previous models, not all parameters in (38) can be identified. Hence we impose the following identifying restrictions on the parameters:[56]

(39) $\omega_{10} = 1$; $\omega_{18} = 1$; $\lambda_{1,1} = 1$; $\lambda_{2,1} = 1$; $\phi_1 = 1$; $\kappa_1 = 1$; $\tau_1 = 1$ and $\mu_1 = 1$.

There are 128 unknown parameters to be estimated. The nonlinear regression model defined by (38) and (39) is our *Model 5*.

As usual, we estimated the unknown parameters for Model 5 using the nonlinear regression option in Shazam. The detailed parameter estimates are listed in the Appendix in Table A9.[57] The R^2 for this model turned out to be 0.8476 and the log likelihood was −8709.9, an increase of 106.0 over the Model 4 log likelihood.[58] Thus adding the 46 extra parameters added significantly to the explanatory power of our hedonic regression model.

When we calculate the price indexes for land in the high and low end wards later in this section, it will be seen that the price movements are quite different, even though the

---

[55] The model defined by equations (38) looks complicated but it is an almost straightforward generalization of Model 4 where we have broken up our observations into two separate groups according to whether the observed sale is in Group 1 or 2 wards. The resulting two Ward models are not completely separate because we force the parameters characterizing the quality adjustment functions $g_A(A_{tn})$, $g_B(B_{tn})$, $f_F(F_{tn})$, $f_M(M_{tn})$ and $f_T(T_{tn})$ to be the same across the two groups of wards.

[56] The restrictions $\lambda_{1,1} = 1$and $\lambda_{2,1} = 1$ replace our old restriction $\lambda_1 = 1$. The other new restriction is $\omega_{18} =1$. Thus the level of land prices in the less expensive wards (the ω_j for j = 5, 12, 15-17 and 19-21) is relative to the level of land prices in Ward 18 where we set $\omega_{18} = 1$. Of course, the movements in land prices in the Group 2 wards is given by the movements in the $\alpha_{2,t}$ and the movements in land prices over time in the Group 1 wards is given by the movements in the $\alpha_{1,t}$. The level of land prices in the Group 1 wards is relative to the level of land prices in Ward 10 where we set $\omega_{10} = 1$.

[57] The standard errors on the estimated high end ward parameters are generally lower (and the T statistics higher) than the estimated lower end ward parameters. There were 3326 observations in the high end wards and only 2252 observations in the lower end wards.

[58] The sum of the residuals in this model was 10.8.

overall land price index has not changed substantially from the land price indexes that resulted from our previous 4 models.

In Model 5, we allow for different schedules of land prices as functions of the plot size in the two types of ward. For high end wards, the relative marginal price of land for small plots is $\lambda_{1,1}$ and this price was set equal to unity. For medium sized plots in high end wards, the marginal price falls to $\lambda_{1,2}$ = 0.8949 but for large sized plots, the marginal price increases to $\lambda_{1,3}$ = 1.0336. For small plots in lower end wards, the relative marginal price of land is $\lambda_{2,1}$ and this price was also set equal to unity. For medium sized plots in lower end wards, the marginal price falls more dramatically to $\lambda_{2,2}$ = 0.6087 but for large sized plots, the marginal price again increases to $\lambda_{2,3}$ = 0.9214. Thus both high and low end wards exhibit the same general pattern of marginal valuations for land as a function of the lot size but the drop in the marginal price is more pronounced for medium sized plots in lower end wards.

Model 5 also allows for different *structure price levels* in high and low end wards. The estimated structure price level parameter for high end wards is β_1 = 3.9734 and for lower end wards, it is β_2 = 2.4777. Thus it appears that the average quality of construction in lower end wards is only about 62% of the construction quality in high end wards.

We turn now to the problems associated with the construction of land, structure and overall price indexes for Tokyo. The construction of the land and overall price indexes is more complex in the present model that in previous models, because the quarter to quarter movements in land prices are different in the Group 1 and 2 wards. For the high end wards, the Model 5 *constant quality residential land price index* for quarter t is defined to be $P_{L1,5t} \equiv \alpha_{1,t}/\alpha_{1,1}$. For the lower end wards, the Model 5 *constant quality residential land price index* for quarter t is defined to be $P_{L2,5t} \equiv \alpha_{2,t}/\alpha_{2,1}$. For all wards, the *constant quality residential structures price index* for quarter t is defined to be the usual MLIT structures price index, $P_{S5t} \equiv p_{Ct}/p_{C1} = p_{Ct}$ since $p_{C1} = 1$. The land and structure price indexes $P_{L1,5t}$, $P_{L2,5t}$ and P_{S5t} for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 5 below.

The Model 5 quarter t *constant quality quantity levels of land in high and lower end wards*, $Q_{L1,5t}$, and $Q_{S2,5t}$ respectively, are defined as the estimated total quarter t values of land in high and lower end wards divided by the corresponding price levels, $P_{L1,5t}$ and $P_{L2,5t}$, for t = 1,...,44:

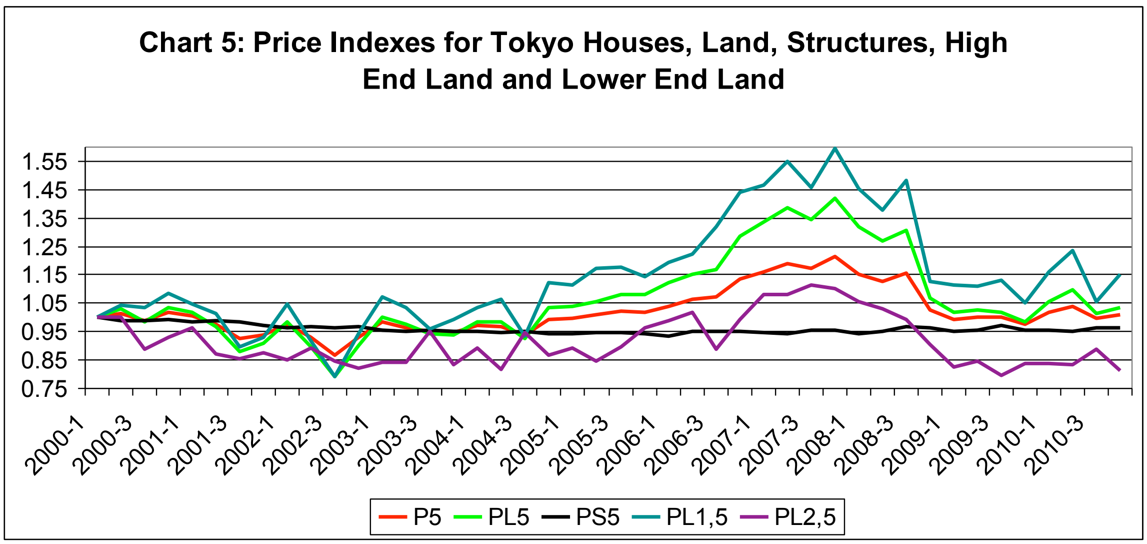
$$(40)\ Q_{L1,5t} \equiv \sum_{n=1}^{N(t)} \alpha_{1,1}\{\omega_1 D_{W,tn,1}+\omega_2 D_{W,tn,2}+\omega_3 D_{W,tn,3}+\omega_4 D_{W,tn,4}+\omega_7 D_{W,tn,7}+\omega_8 D_{W,tn,8} +\omega_9 D_{W,tn,9}+\omega_{10} D_{W,tn,10}+\omega_{11} D_{W,tn,11}+\omega_{13} D_{W,tn,13}+\omega_{14} D_{W,tn,14}\} f_{1L}(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn});$$

$$(41)\ Q_{L2,5t} \equiv \sum_{n=1}^{N(t)} +\alpha_{2,1}\{\omega_5 D_{W,tn,5}+\omega_6 D_{W,tn,6}+\omega_{12} D_{W,tn,12}+\omega_{15} D_{W,tn,15}+\omega_{16} D_{W,tn,16} +\omega_{17} D_{W,tn,17}+\omega_{18} D_{W,tn,18}+\omega_{19} D_{W,tn,19}+\omega_{20} D_{W,tn,11}+\omega_{21} D_{W,tn,21}\} f_{2L}(L_{tn}) f_F(F_{tn}) f_M(M_{tn}) f_T(T_{tn}).$$

The Model 5 quarter t *constant quality quantity structure level*, Q_{S5t}, is defined as the total quarter t estimated value of structures divided by the corresponding price level $P_{S5t} = p_{Ct}$ for $t = 1,...,44$:[59]

$$(42)\ Q_{S5t} \equiv \sum_{n=1}^{N(t)} \beta_1\{D_{W,tn,1}+D_{W,tn,2}+D_{W,tn,3}+D_{W,tn,4}+D_{W,tn,7}+D_{W,tn,8}+D_{W,tn,9}+D_{W,tn,10} + D_{W,tn,11}+D_{W,tn,13}+D_{W,tn,14}\}g_A(A_{tn})g_B(B_{tn})S_{tn} + \sum_{n=1}^{N(t)} \beta_2\{D_{W,tn,5}+D_{W,tn,6}+D_{W,tn,12}+D_{W,tn,15}+D_{W,tn,16}+D_{W,tn,17}+D_{W,tn,18} + D_{W,tn,19}+D_{W,tn,20}+D_{W,tn,21}\}g_A(A_{tn})g_B(B_{tn})S_{tn}.$$

We use the Fisher ideal index to aggregate the price and quantity components for high and lower end land. Thus define the *overall land price level for quarter t* for Model 5, P_{L5t}, as the chained Fisher price index of the two land price and quantity series $\{P_{L1,5t},P_{L2,5t},Q_{L1,5t},Q_{L2,5t}\}$. The *overall house price index for Tokyo for quarter t* for Model 5, P_{5t}, is defined as the chained Fisher price index of the two land price and quantity series and the structure price and quantity series, $\{P_{L1,5t},P_{L2,5t},P_{S5t},Q_{L1,5t},Q_{L2,5t},Q_{S5t}\}$. The overall Model 5 house price index P_{5t} as well as the overall land price index P_{L5t} for Tokyo over the 44 quarters in the years 2000-2010 are also graphed in Chart 5 below.[60]



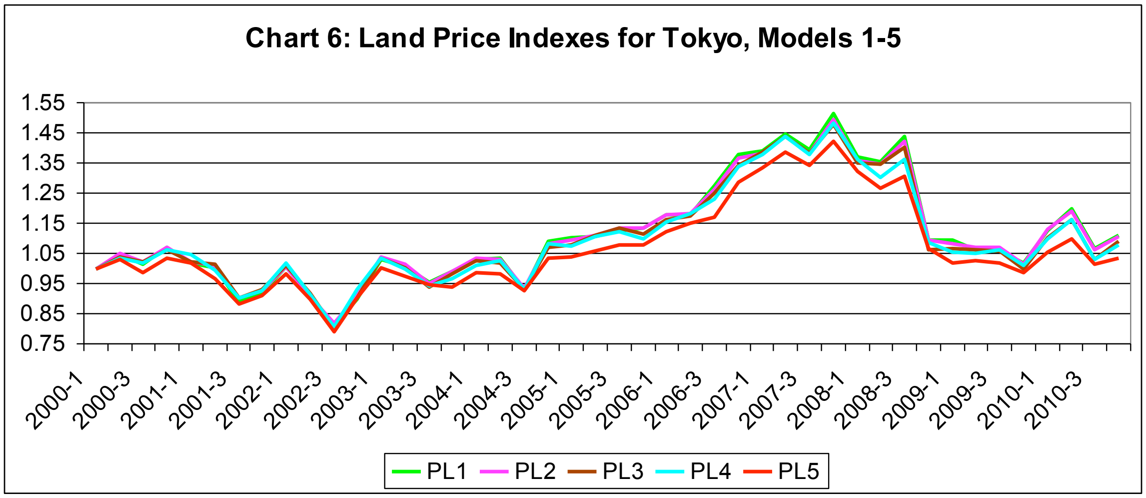
As expected, the pattern of land price movements is very different in the high and low end wards. Price movements have generally been higher and more volatile in the more expensive wards; i.e., $P_{L1,5t}$ generally lies above $P_{L2,5t}$ and $P_{L1,5t}$ has a higher variance than $P_{L2,5t}$.[61] However, it can also be seen from viewing Chart 5 that the overall land price index for Model 5, P_{L5t}, is not that different from the land price indexes from previous Models.[62] We compare the Model 1 to Model 5 overall land price indexes in Chart 6.

[59] Note that p_{Ct} does not appear on the right hand side of (42).

[60] The series P_{5t}, P_{L5t}, $P_{L1,5t}$ and $P_{L2,5t}$ are listed in Table A10 of the Appendix. P_{S5t} is equal to P_{S4t} and is listed in Table A8 and is equal to the MLIT construction index p_{Ct}.

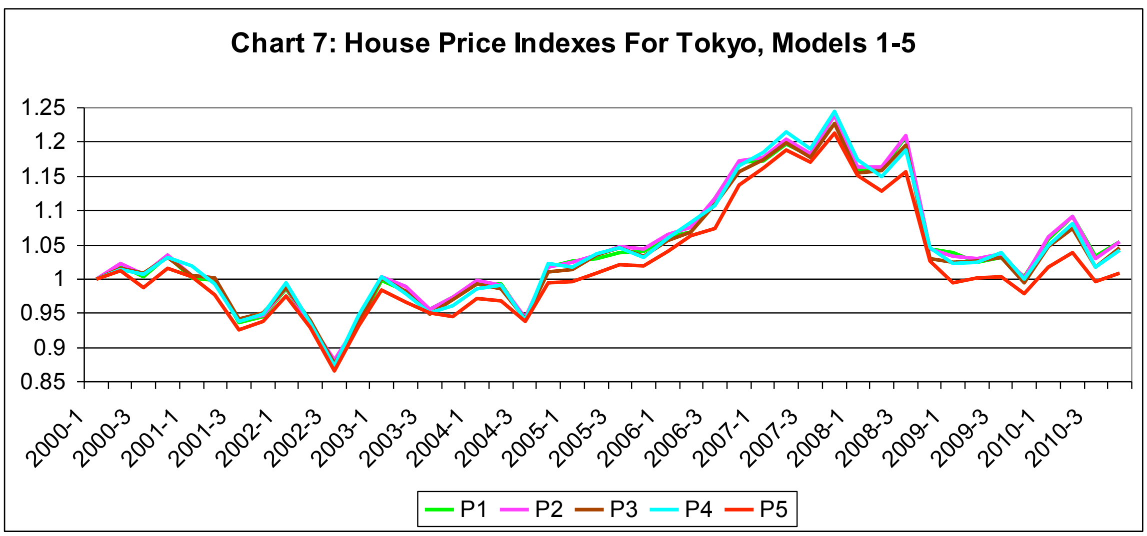
[61] The sample variance for $P_{L1,5t}$ was 0.0358 and for $P_{L2,5t}$ was 0.0077.

[62] The correlation coefficients between P_{L5} and P_{L1}, P_{L2}, P_{L3} and P_{L4} were 0.98997, 0.99123, 0.99324 and 0.99684 respectively.



It can be seen that the overall land price series for Models 1-4, P_{L1t}-P_{L4t}, are generally quite close with small drops in the series as we move from Model 1 to Model 4. The overall land price series for Model 5 drops a more substantial amount: about 3% on average.[63] However, the overall pattern of land price movements is much the same in all 5 Models.

It is also useful to compare the overall house price indexes for Models 1-5 and this is done in Chart 7.



Again, there are only small differences in the overall house price indexes P_{1t}-P_{4t} for Models 1-4.[64] However, the Model 5 overall house price index P_{5t} is about 2% lower on average compared to the levels in the other Models.

[63] We regard the Model 5 estimates as the most accurate estimates since this model has reasonable parameter values and gives us the best fit.

[64] The correlation coefficients between P_5 and P_1, P_2, P_3 and P_4 were 0.98637, 0.98822, 0.99132 and 0.99492.

In summary, the overall house price index P_{5t} is probably the most accurate one but the overall pattern of price movements is much the same in all 5 Models. Two important implications of our results for statistical agencies are as follows:

- Our generalized builder's model can provide a sensible decomposition of house prices in a major city into land and structure components and
- Model 1, our simplest model that uses only information on lot size, floor space size, the age of the structure and the ward in which the lot is located, can provide an adequate approximation to a more data intensive model that uses information on other characteristics of the lot location and the structure.

All of the price indexes that we have constructed thus far have been for the quarterly *sales* of houses in Tokyo. In order to construct estimates of real household wealth, it is useful to be able to construct price indexes for the *stock* of residential houses in Tokyo. In the following section, we show how approximate stock indexes can be constructed using the Models that have already been estimated.

8. Approximate Stock House Price and Land Price Indexes

In order to construct a completely accurate price index for the stock of houses in a city or location, it is necessary to have an updated census of dwelling units in the area under consideration. However, if census information is not available, it is possible to construct an approximation to a housing stock price index for the location using cumulated information on the sales of houses in the location.[65]

The basic idea is straightforward: we form an approximation to the quantity of quality adjusted *high end land*, *lower end land* and *structures* over our sample period by cumulating the corresponding quarterly sales quantities $Q_{L1,5t}$, $Q_{L2,5t}$ and Q_{S5t} defined by equations (40)-(42) in the previous section. Define the *cumulated quantities* as follows:[66]

$$(43)\ Q_{L1} \equiv \sum_{t=1}^{44} Q_{L1,5t};\ Q_{L2} \equiv \sum_{t=1}^{44} Q_{L2,5t};\ Q_S \equiv \sum_{t=1}^{44} Q_{S5t}.$$

The corresponding land prices are the Model 5 land prices defined in the previous section: for the higher and lower end wards, the *constant quality residential approximate stock land price indexes* for quarter t are defined to be $P_{L1,t} \equiv \alpha_{1,t}/\alpha_{1,1}$ and $P_{L2,t} \equiv \alpha_{2,t}/\alpha_{2,1}$. The *constant quality residential structures stock price index* for quarter t is defined to be the usual MLIT structures price index, $P_{St} \equiv p_{Ct}$.

Our *approximate land price for the stock of houses* in Tokyo for quarter t, P_{KL5t}, based on the Model 5 regression parameters is defined as the following Lowe (1823) index:[67]

[65] This approximate stock of housing price index methodology was explained in Chapter 8 of the Eurostat *Residential Property Price Index Handbook*; see de Haan and Diewert (2011; sections 8.49-8.57).
[66] These cumulated quantities divided by the sample number of observations 5578 turned out to equal Q_{L1}= 2.22644, Q_{L2} = 1.19465 and Q_S = 2.73970.
[67] For additional material on Lowe indexes, see Hill (2004; Ch. 15).

(44) $P_{KL5t} \equiv [P_{L1,t}Q_{L1}+P_{L2,t}Q_{L2}]/[P_{L1,1}Q_{L1}+P_{L2,1}Q_{L2}]$; $t = 1,...,44.$

It can be seen that the land price index defined by (44) is a fixed basket type index where the quantity basket consists of the quality adjusted total amounts of the two types of residential land in Tokyo.[68]

Our *approximate overall house price for the stock of houses* in Tokyo for quarter t, P_{K5t}, based on the Model 5 regression parameters is defined as the following Lowe index:

(45) $P_{K5t} \equiv [P_{L1,t}Q_{L1}+P_{L2,t}Q_{L2}+P_{St}Q_{St}]/[P_{L1,1}Q_{L1}+P_{L2,1}Q_{L2}+P_{St}Q_{S1}]$; $t = 1,...,44.$

The land and overall stock price indexes P_{K5t} and P_{KL5t} defined by (44) and (45) are compared with their sales counterparts from Model 5, P_{5t} and P_{L5t}, in Chart 8.[69]

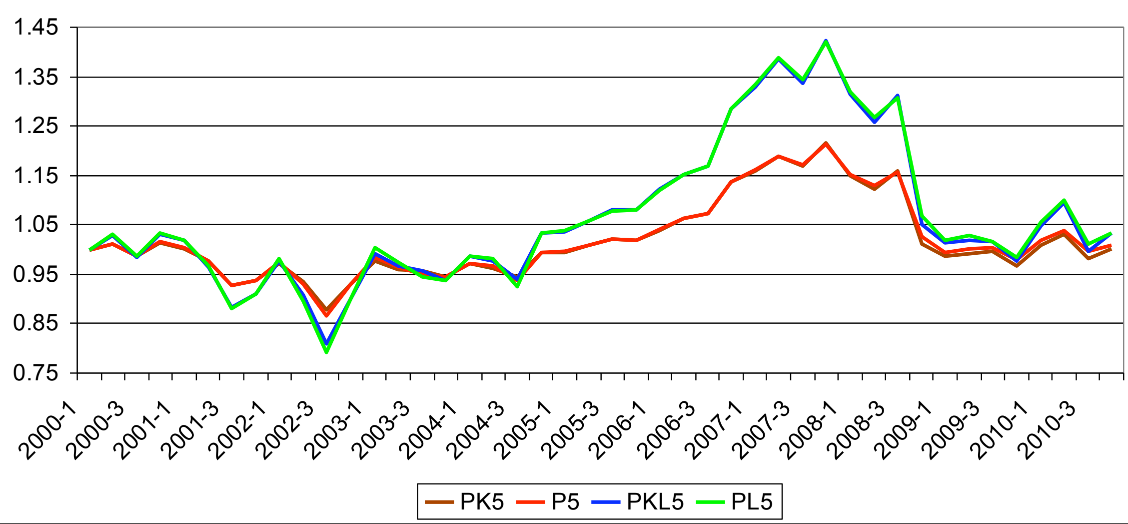


Chart 8: Approximate Stock and Sales House and Land Price Indexes Based on Model 5 Estimates

It can be seen that the overall approximate housing stock index P_{K5t} is very close to its sales counterpart P_{5t} and the approximate stock of land price index P_{KL5t} is almost identical to its sales counterpart P_{L5t}.[70] These close correspondences are very encouraging since it indicates that the sales based indexes are likely to provide adequate

[68] When quantities are constant across periods as they are in the case of a Lowe index, it will turn out that fixed base and chained Laspeyres, Paasche and Fisher indexes will all be equal.

[69] The corresponding series are listed in Table A10 in the Appendix.

[70] The differences in the stock type indexes and their sales counterparts are entirely due to the effects of different quantity weights since the price components are identical in these counterpart indexes. The close correspondence of P_{K5t} to P_{5t} shows that the quarter to quarter fluctuations in $Q_{L1,5t}$, $Q_{L2,5t}$ and Q_{S5t} (compared to the fixed weights Q_{L1}, Q_{L2} and Q_S) were not large enough to cause the stock and sales type indexes to diverge substantially.

approximations to the corresponding true stock indexes, which use updated census weights for the housing stock.

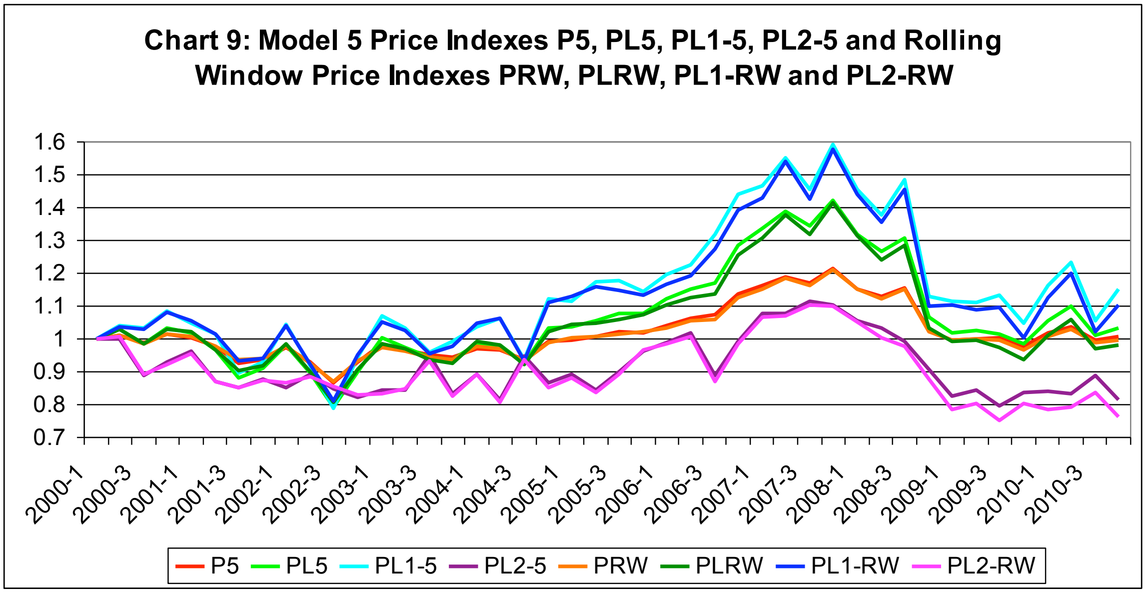
The sales price hedonic regression models that we have presented in previous sections are not completely suitable for use by statistical agencies producing house price indexes. The reason for this lack of suitability is due to the fact that as data on sales for the most recent quarter becomes available, the new hedonic regression will give rise to new estimates of house price inflation for past periods and this would lead to a need to revise past series. For many purposes, it is useful to have price indexes that are not revised. In the following section, we will address how to deal with this revisions problem.

9. Rolling Window Hedonic Housing Regressions

We dealt with the no revisions problem in the following way. We started off by using Model 5 but applied it to only the first 24 quarters of our sample (instead of the full 44 quarters). We then computed our land, structures and overall house price indexes as in section 7 above for quarters 1-24. At Stage 2 of our procedure, we dropped the data for quarter 1 and added the data for quarter 25 to form our Stage 2 data set and then ran the nonlinear regression model defined by equations (38) and (39) for quarters t = 2,3,...,25. Using these new coefficient estimates, we computed the structure price index and land price indexes for high and low end wards as in section 7 for quarters 2-25. However, we used only the *ratios* of the Stage 2 quarter 25 to quarter 24 land price indexes in order to *update* our previous Stage 1 land price indexes so that the new set of indexes covered quarters 1-25.[71] At Stage 3 of our procedure, we dropped the data for quarter 2 and added the data for quarter 26 from our Stage 2 data to form our Stage 3 data set and then ran the nonlinear regression model defined by equations (38) and (39) for quarters t = 3,4,...,26. Using these new coefficient estimates, we computed the structure price index and land price indexes for high and low end wards as in section 7 for quarters 3-26. We used only the ratios of the Stage 3 quarter 26 to quarter 25 land price indexes in order to update our previous Stage 2 land price indexes so that the new set of indexes covered quarters 1-26. We continued this process of adding the data of the next quarter and dropping the data of the oldest quarter in the *rolling window* of 24 quarters until we reached quarter 44. Thus we ran a total of 21 *Rolling Window Hedonic Regressions*.[72] The resulting Rolling Window overall house price indexes P_{RWt}, overall land price indexes P_{LRWt}, high and low end ward land price indexes, P_{L1RWt} and P_{L2RWt}, are plotted in Chart 9[73] along with their Model 5 counterpart indexes, P_{5t}, P_{L5t}, $P_{L1,5t}$ and $P_{L2,5t}$.

---

[71] The structure price indexes turn out to equal the MLIT indexes p_{Ct} that we have listed previously.
[72] Each of the 21 regressions had 88 parameters to estimate with a varying number of degrees of freedom. This rolling window updating procedure was introduced by Shimizu, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010) in their hedonic regression models for Tokyo house prices. The method we are using here deals with the extra complications due to the need for separate land and structures estimates. Our present method was explained and implemented for the Dutch town of "A" with a window length of 9 quarters; see Chapter 8 of de Haan and Diewert (2011). The rolling window updating methodology has also been used previously in an index number context; see Ivancic, Diewert and Fox (2011), de Haan and van der Grient (2011) and de Haan and Krsinich (2012).
[73] These indexes are also listed in Table A12 in the Appendix.



Viewing Chart 9, it can be seen that the Model 5 overall house price index, P_5, can hardly be distinguished from its Rolling Window counterpart, P_{RW}. However, for the land price indexes, it can be seen that while the Model 5 indexes P_{L5} (the overall land price index), $P_{L1,5}$ (the high end ward land price index) and $P_{L2,5}$ (the lower end ward land price index) are very close to their Rolling Window counterparts P_{LRW}, P_{L1RW} and P_{L2RW} for the first 5 years in our sample, the Rolling Window land price indexes tend to be *lower* than their single regression Model 5 counterparts for the last 5 years in our sample.

The question now arises: which model should be a preferred model: Model 5 based on a single regression or the Rolling Window Model based on 21 separate hedonic regressions? We prefer the Rolling Window Model since it allows for gradual change in the hedonic coefficients over time and moreover, the RW Model fits the data better while still generating sensible parameter estimates.[74]

Our conclusion here is that the Rolling Window hedonic house price regression model is a suitable one for a statistical agency that is mandated to produce a house price index in a timely manner without having to make revisions to previous estimates. An open question which we did not explore in the present paper is the question of choosing the "optimal" window length.

10. Comparison with Traditional Time Dummy Hedonic Regression Models

There is no doubt that our Model 5 defined in section 7 is rather complicated. Thus most hedonic housing regression models are based on the far simpler *time dummy approach*

[74] The R^2 for the 21 regressions were as follows: 0.8518, 0.8470, 0.8494, 0.8481, 0.8496, 0.8489, 0.8526, 0.8511, 0.8515, 0.8483, 0.8505, 0.8533, 0.8551, 0.8544, 0.8540, 0.8551, 0.8551, 0.8525, 0.8516, 0.8555 and 0.8573. Recall that the R^2 for the Model 5 regression was 0.8476. The structural parameters that were common to all of the regressions did not change much over time but there were some small changes which of course led to the differences between the Model 5 and Rolling Window land price indexes.

where the log of the selling price of the house is regressed on either a linear function of the characteristics or on the logs of the characteristics of the house along with time dummy variables.[75] This method does not generate decompositions of the selling price into land and structure components and so it is not suitable when such decompositions are required but the time dummy method can be used to generate overall house price indexes. In this section, we will use the time dummy method to generate overall house price indexes and compare them with our Model 5 overall estimates.

Recall that V_{tn} is the sales price of property n that was sold in quarter t, L_{tn} is the area of the plot, S_{tn} is the floor space area of the structure and A_{tn} is the age of the structure. In the time dummy linear regression defined below by (46), we have replaced V_{tn}, L_{tn}, S_{tn} and A_{tn} by their logarithms, $\ln V_{tn}$, $\ln L_{tn}$, $\ln S_{tn}$ and $\ln A_{tn}$.[76] Our first time dummy hedonic regression model is defined for t = 1,...,44 and n = 1,...,N(t) by the following equations:

$$(46)\ \ln V_{tn} = \alpha_t + \gamma \ln L_{tn} + \beta \ln S_{tn} + \delta \ln A_{tn} + \sum_{j=1}^{21} \omega_j D_{W,tn,j} + \varepsilon_{tn}$$

where $\alpha_1,...,\alpha_{44}$, γ, β, δ and $\omega_1,...,\omega_{21}$ are 68 unknown parameters to be estimated and $D_{W,tn,j}$ is the Ward j dummy variable for observation tn defined earlier by (4). The α_t are the quarter t time coefficients which shift the hedonic surface during each quarter, γ and β are parameters which adjusts the sales price for the size of the lot and the floor space area respectively, δ is a parameter which adjusts the sales price for the age of the structure (essentially a depreciation parameter) and the ω_j are parameters which adjust the selling price V_{tn} up or down depending on the ward that property n in quarter t is located. We expect β and γ to be positive and δ to be negative. The time dummy variables associated with the α_t and the dummy variables $D_{W,tn,j}$ associated with the wards are linearly dependent and so we need to impose a normalization on the parameters in order to identify the remaining parameters. We choose the following normalization:

$$(47)\ \alpha_1 = 0.$$

Model 6 is the hedonic regression model defined by (46) and (47). Using our Tokyo housing data, we estimated Model 6 using the Ordinary Least Squares option in Shazam and the parameter estimates are listed in Table A13 in the Appendix.[77] The R^2 for this model turned out to be 0.8432 with a log likelihood of 1985.9.

---

[75] This methodology was developed by Court (1939; 109-111) as his Hedonic Suggestion Number Two. For an application of the time dummy approach to the construction of house price indexes for Tokyo, see Shimizu and Nishimura (2007).

[76] The log-linear regression model that replaced $\ln L_{tn}$, $\ln S_{tn}$ and $\ln A_{tn}$ by their levels, L_{tn}, S_{tn} and A_{tn}, led to an R^2 of 0.8374 and a log likelihood of 1883.5 which is lower than the R^2 and log likelihood generated by the model defined by (46). Thus we report only the results for the log-log model.

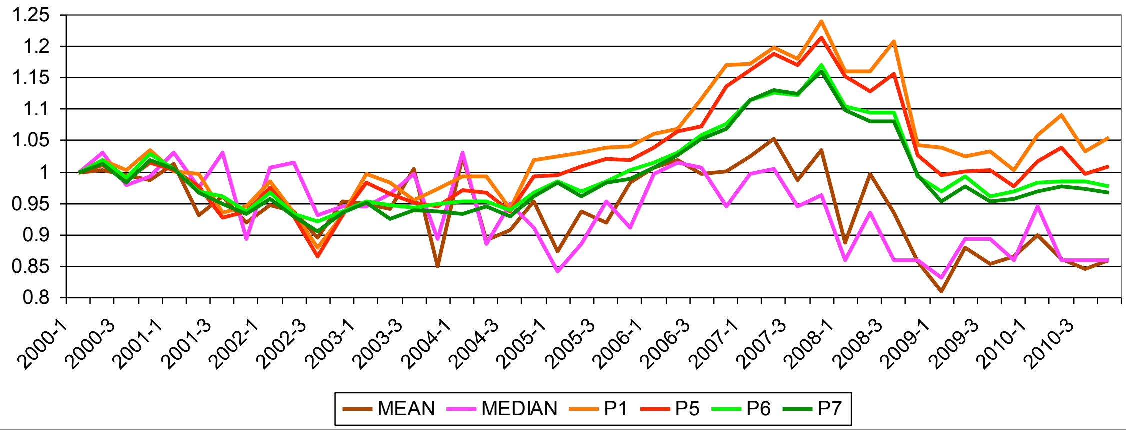
[77] The estimated structure and land area coefficients turned out to be $\beta = 0.44108$ and $\gamma = 0.49708$ with T statistics of 40.32 and 56.01. The estimated age of structure parameter turned out to be $\delta = -0.09662$ with a T statistic of -27.70. The coefficients ω_j associated with the Ward dummy variables were highly significant with T statistics ranging between 43.6 and 131.8. For Model 5, the expensive Wards were 1-4, 7-11 and 13-14. For the present Model 6, the Wards with the highest ω_j's were 1-4 and 7-14. The sum of residuals was 0 in Models 6 and 7.

The overall house price indexes for Model 6, P_{6t}, are defined as the exponentials of the time coefficients α_t:

(48) $P_{6t} \equiv \exp[\alpha_t]$; $t = 1,....,44.$

The Model 6 time dummy hedonic regression uses the same characteristics information that we used in our nonlinear regression for Model 1. Thus we compare our Model 6 overall price indexes P_{6t} with the Model 1 indexes P_{1t} and our best Model 5 indexes P_{5t} in Chart 10 below[78] along with the Mean and Median indexes.

Chart 10: Mean, Median, Nonlinear Hedonic Models P1 and P5 and Time Dummy Hedonic Models P6 and P7 House Price Indexes for Tokyo



It can be seen that during the last half of the sample period, the Mean and Median house price series are about 10-20% below the other indexes P_{1t}, P_{5t} and P_{6t} that rely on hedonic regressions to control for the quality of the houses sold in each quarter. Over the entire sample period, the time dummy index P_{6t} is on average about 2% below our best nonlinear regression based index P_{5t} and about 4% below our initial nonlinear regression based index P_{1t} that used the same characteristics information that was used in Model 6. Thus the time dummy based indexes P_{6t} do differ somewhat from our overall house price indexes P_{1t}-P_{5t} that were based on variants of our basic builder's model.

We run one additional time dummy model that uses all of the characteristics information that we used in Model 5. Perhaps the use of this extra information will lead to an index which is close to P_{5t}.

We will add the following 4 variables as explanatory variables to the regression model defined earlier by (46): NB = Number of bedrooms; WI = Width of the lot in meters; TW = Walking time in minutes to the nearest subway station and TT = Subway running time

[78] The P_{6t} are also listed in Table A14 in the Appendix.

in minutes to the Tokyo station from the nearest station.[79] Thus our *Model 7* regression is defined as follows, for t = 1,...,44 and n = 1,...,N(t):

(49) $\ln V_{tn} = \alpha_t+\gamma \ln L_{tn}+\beta \ln S_{tn}+\delta \ln A_{tn}+\phi NB_{tn}+\kappa \ln WI_{tn}+\tau TW_{tn}+\mu TT_{tn}+\sum_{j=1}^{21} \omega_j D_{W,tn,j}+\varepsilon_{tn}$

where $\alpha_1,...,\alpha_{44}$, γ, β, δ, ϕ, κ, τ, μ and $\omega_1,...,\omega_{21}$ are 72 unknown parameters to be estimated and $D_{W,tn,j}$ is the Ward j dummy variable for observation tn defined earlier by (4). As was the case with Model 6, not all parameters are identified. Thus we again impose the normalization (47), which was $\alpha_1 = 0$. The 4 new parameters ϕ, κ, τ and μ are associated with the variables NB_{tn}, WI_{tn}, TW_{tn} and TT_{tn}. We tried entering each of these variables into the regression defined by (49) in levels form or by transforming the variable by the natural logarithm function. We found that entering the lot width variable in log form led to a higher log likelihood (so notice that we have the term $\kappa \ln WI_{tn}$ in (49) rather than the term κWI_{tn}) but for the other 3 new variables, the levels form led to higher log likelihoods. Using our Tokyo housing data, we estimated Model 7 using the Ordinary Least Squares option in Shazam and the parameter estimates are listed in Table A15 in the Appendix.[80] The R^2 for this model turned out to be 0.8621 with a log likelihood of 2344.1, a very large increase of 358.3 over the Model 6 log likelihood.

The overall house price indexes for Model 7, P_{7t}, are defined as the exponentials of the new time parameters α_t:

(50) $P_{7t} \equiv \exp[\alpha_t]$; $\qquad$ t = 1,....,44.

The Model 7 time dummy hedonic regression uses the same characteristics information that we used in our nonlinear regression for Model 5. Thus we compared our Model 7 overall price indexes P_{7t} with the Model 5 indexes P_{5t} in Chart 10 above[81] along with the Mean and Median indexes. It can be seen that the P_{7t} are not all that different from the P_{6t} and both indexes are still generally below our best index P_{5t}.[82]

Viewing Chart 10, it can be seen that the time dummy regression models give rise to house price indexes P_{6t} and P_{7t} that are fairly close to our preferred index P_{5t} but there are two significant differences:

---

[79] These variables were defined in section 2 above.

[80] The estimated structure and land area coefficients turned out to be $\beta = 0.42882$ and $\gamma = 0.52920$ with T statistics of 38.7 and 62.7. The estimated age of structure parameter turned out to be $\delta = -0.08885$ with a T statistic of −26.5. The estimated width, bedrooms, walking time and subway time parameters were $\kappa = 0.10277$, $\phi = -0.00190$, $\tau = -0.00106$ and $\mu = -0.00007$ with T statistics of 11.4, −7.4, −20.6 and −16.9. The coefficients ω_j associated with the Ward dummy variables were highly significant with T statistics ranging between 45.2 and 92.4. For Model 5, the expensive Wards were 1-4, 7-11 and 13-14. For the present Model 7, the expensive Wards with the highest ω_j's were 1-4 and 7-14. The signs of the estimated parameters are all reasonable.

[81] The P_{7t} are also listed in Table A14 in the Appendix.

[82] The sample means (over the 44 quarters) for the P_{1t}, P_{5t}, P_{6t} and P_{7t} were 1.0404, 1.0199, 1.0010 and 0.9935 respectively. Thus on average, the P_{7t} were 0.75% below the corresponding P_{6t} and 2.64% below the P_{5t}. The correlation coefficients between P_5 and P_1, P_6, P_7 were 0.98637, 0.96770, 0.96507 respectively.

- The P_{6t} and P_{7t} are significantly *below* the corresponding P_{5t} and
- The P_{6t} and P_{7t} are significantly *smoother* than the corresponding P_{5t}.[83]

Thus the time dummy based house price indexes do differ significantly from the indexes generated by our best builder's model. However, can we determine which type of model is "best"? In order to answer this question, we return to the basic builder's model defined by equations (1) above, which set the value of a property, V_{tn}, equal to the sum of its land value, $\alpha_t L_{tn}$, plus its structure value, $\beta_t S_{tn}$. Thus there are *four main determinants of property value* in this simplified model: the land area L_{tn}, the structure area S_{tn}, the period t price of land α_t and the period t price of the structure per meter squared β_t. The corresponding simplified time dummy hedonic regression model sets the logarithm of property value, $\ln V_{tn}$, equal to a time dummy parameter, say ρ_t, plus a weighted sum of the logarithms of land and structure areas, $\theta \ln L_{tn} + \sigma \ln S_{tn}$. Thus the exponential of ρ_t, say $\pi_t \equiv \exp[\rho_t]$, can be interpreted as the *period t price of the fixed weight "average"*[84] *of land and structure composite commodity*, $L_{tn}{}^{\theta} S_{tn}{}^{\sigma}$. The weakness of the time dummy model now becomes apparent: the time dummy model has only a *single price* π_t for a fixed weight aggregate of land and structures that can vary over time whereas the builder's model has *two prices* that can vary independently over time, the prices of land and structures, α_t and β_t. Thus at this stage of the argument, it is clear that the builder's model is a far more flexible model than the time dummy model.[85] However, in our empirical work, we did not estimate the movements of structure prices over time; i.e., we assumed that an official house construction price index could accurately capture how the price of structures changed over time.[86] If this assumption is far from being satisfied, then it is possible that the time dummy model could give more accurate results. Our subjective assessment is that the MLIT construction price index does reflect movements in house construction costs in Tokyo and thus we feel that the Model 5 results are more accurate than the Model 6 and 7 results. However, all three models provide similar overall price indexes for house sales in Tokyo.

11. Conclusion

We summarize some of the main points that have emerged in the previous sections:

- The paper shows that the builder's model that was previously applied to a small Dutch town[87] can be applied to a large urban city (Tokyo) provided that we have information on the sales price of houses, the land and structure areas of the house, the age of the house, some information on the location of the properties and an exogenous price index for house construction costs. The builder's model can

[83] The sample variances for the P_{1t}, P_{5t}, P_{6t} and P_{7t} are 0.00692, 0.00624, 0.00378 and 0.00396.
[84] We have a true average only if θ and σ sum to one.
[85] The two models can give the same answer empirically if either $\alpha_t = \lambda\beta_t$ for all periods t so that the prices of land and structures move proportionally over time or if $L_{tn} = \mu S_{tn}$ for all t and n so that the land-structure ratio is constant for all properties. Neither possibility is empirically likely.
[86] Recall that we made this assumption to eliminate the multicollinearity problem between L_{tn} and S_{tn}.
[87] See de Haan and Diewert (2011) and Diewert, de Haan and Hendriks (2011a) (2011b).

successfully provide a decomposition of property value into land and structures components.

- We showed how additional information on the characteristics of the properties can be incorporated into the builder's model, leading to models that fit the data better and thus presumably providing more accurate land, structure and overall price indexes.
- Hedonic regression models typically model the effects of increasing amounts of a characteristic on the selling price in a linear fashion. In the present paper, we generalized this approach to allow the response to be a piece-wise linear function (or spline function) in place of a linear response function. This generalization was particularly important in modeling the effects of structure age and of walking time to the nearest subway station; see sections 4 and 6.
- In section 8, we showed how our regression results could be used to calculate approximate price indexes for the *stock* of houses in Tokyo (as opposed to the *sales* of houses in each quarter).
- In section 9, we computed 21 Rolling Window regressions and showed how these regression results could be used to construct house price indexes that were timely and did not need to be revised each period as new information on house sales became available.
- Finally, in section 10, we compared our builder's model indexes to traditional time dummy hedonic regression models. The comparisons could only be made for the overall house price indexes since the time dummy method does not lead to accurate separate indexes for land and for structures. We found that while the time dummy and builder's models captured the same trends, there were some small but significant differences between the indexes generated by the two approaches.

Appendix: Model Estimated Coefficients and Index Number Tables

Table A1: Estimated Coefficients for Model 1

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
ω_1	2.1348	41.112	α_3	3.7863	28.383	α_{25}	4.4053	35.093
ω_2	1.0020	30.511	α_4	3.9980	32.103	α_{26}	4.3998	35.979
ω_3	1.1553	30.269	α_5	3.7944	32.603	α_{27}	4.7558	31.124
ω_4	1.0552	11.541	α_6	3.7475	27.506	α_{28}	5.1506	40.423
ω_5	0.38569	5.621	α_7	3.3218	26.688	α_{29}	5.1939	37.356
ω_6	0.62467	9.992	α_8	3.4285	30.338	α_{30}	5.4013	37.140
ω_7	1.0214	27.35	α_9	3.7525	27.488	α_{31}	5.2080	33.905
ω_8	1.2304	58.353	α_{10}	3.3802	28.813	α_{32}	5.6581	39.967
ω_9	0.88449	46.691	α_{11}	3.0205	23.868	α_{33}	5.1146	31.804
ω_{11}	1.6639	41.882	α_{12}	3.3602	31.929	α_{34}	5.0592	31.877
ω_{12}	0.67269	34.870	α_{13}	3.8478	29.689	α_{35}	5.3721	32.813
ω_{13}	0.79505	64.468	α_{14}	3.7603	32.321	α_{36}	4.0782	23.219
ω_{14}	0.89487	26.294	α_{15}	3.5570	28.634	α_{37}	4.0863	22.016
ω_{15}	0.54123	8.8738	α_{16}	3.7025	22.845	α_{38}	3.9651	24.827
ω_{16}	0.44453	6.0919	α_{17}	3.8440	34.010	α_{39}	3.9528	24.771
ω_{17}	0.45904	16.009	α_{18}	3.8632	29.935	α_{40}	3.8021	23.690
ω_{18}	0.49218	39.188	α_{19}	3.4764	28.183	α_{41}	4.2077	27.508
ω_{19}	0.21120	8.9117	α_{20}	4.0631	30.474	α_{42}	4.4752	28.542
ω_{20}	0.28298	7.9508	α_{21}	4.1170	31.375	α_{43}	3.9829	25.538

ω_{21}	0.33419	12.273	α_{22}	4.1321	31.351	α_{44}	4.1515	29.487
α_1	3.7342	32.491	α_{23}	4.1994	28.264	β	3.4071	59.780
α_2	3.9089	33.202	α_{24}	4.2315	35.553	δ	0.01394	26.830

Table A2: Mean and Median House Price Indexes for Tokyo, Model 1 Overall Price Index P_1, Land Price Index P_{L1} and Structure Price Index P_{S1}

Quarter	P_{Mean}	P_{Median}	P_1	P_{L1}	P_{S1}
2000-1	1.00000	1.00000	1.00000	1.00000	1.00000
2000-2	1.00349	1.03192	1.01926	1.04678	0.98919
2000-3	0.99552	0.98016	1.00253	1.01395	0.98919
2000-4	0.98649	0.99223	1.03399	1.07064	0.99459
2001-1	1.01299	1.03192	1.00091	1.01612	0.98378
2001-2	0.93072	0.97498	0.99714	1.00355	0.98919
2001-3	0.96159	1.03192	0.93555	0.88956	0.98378
2001-4	0.91955	0.89387	0.94532	0.91814	0.97297
2002-1	0.94738	1.00690	0.98594	1.00490	0.96216
2002-2	0.93671	1.01467	0.93521	0.90521	0.96757
2002-3	0.89508	0.93184	0.88051	0.80889	0.96216
2002-4	0.95421	0.94564	0.93233	0.89983	0.96757
2003-1	0.94934	0.94564	0.99763	1.03042	0.95676
2003-2	0.94085	0.96462	0.98258	1.00700	0.95135
2003-3	1.00603	0.99741	0.95549	0.95256	0.95676
2003-4	0.85028	0.89387	0.97334	0.99152	0.95135
2004-1	1.02468	1.03192	0.99288	1.02941	0.95135
2004-2	0.89192	0.88525	0.99299	1.03454	0.94595
2004-3	0.90729	0.94909	0.94139	0.93095	0.95135
2004-4	0.95412	0.91113	1.01828	1.08808	0.94054
2005-1	0.87366	0.84211	1.02551	1.10250	0.94054
2005-2	0.93691	0.88525	1.03026	1.10654	0.94595
2005-3	0.91959	0.95427	1.03932	1.12457	0.94595
2005-4	0.98333	0.91113	1.04097	1.13316	0.94054
2006-1	1.00718	0.99741	1.06178	1.17972	0.93514
2006-2	1.01915	1.01467	1.06910	1.17824	0.95135
2006-3	0.99796	1.00777	1.11739	1.27359	0.95135
2006-4	1.00189	0.94564	1.16990	1.37930	0.95135
2007-1	1.02574	0.99741	1.17282	1.39090	0.94595
2007-2	1.05370	1.00604	1.19774	1.44645	0.94054
2007-3	0.98805	0.94564	1.18005	1.39467	0.95676
2007-4	1.03498	0.96290	1.23907	1.51522	0.95676
2008-1	0.88688	0.85936	1.16089	1.36966	0.94054
2008-2	0.99726	0.93615	1.15999	1.35483	0.95135
2008-3	0.93450	0.85936	1.20819	1.43862	0.96757
2008-4	0.85835	0.85936	1.04315	1.09213	0.96216
2009-1	0.80939	0.83261	1.03838	1.09430	0.95135
2009-2	0.88014	0.89387	1.02507	1.06183	0.95676
2009-3	0.85473	0.89387	1.03207	1.05855	0.97297
2009-4	0.86533	0.85936	1.00357	1.01819	0.95676
2010-1	0.90025	0.94564	1.05875	1.12679	0.95676
2010-2	0.86150	0.85936	1.09112	1.19843	0.95135
2010-3	0.84646	0.85936	1.03284	1.06660	0.96216
2010-4	0.85947	0.85936	1.05487	1.11175	0.96216

Table A3: Estimated Coefficients for Model 2

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
ω_1	2.0767	40.384	α_5	4.1769	28.529	α_{28}	5.573	35.504
ω_2	0.9913	37.299	α_6	4.1243	6.404	α_{29}	5.6444	33.495
ω_3	1.1570	33.140	α_7	3.6852	23.834	α_{30}	5.8674	33.101
ω_4	1.0095	11.842	α_8	3.7734	26.696	α_{31}	5.6407	15.302
ω_5	0.3983	7.420	α_9	4.1041	22.655	α_{32}	6.0943	31.948
ω_6	0.6319	12.675	α_{10}	3.6993	26.211	α_{33}	5.5414	27.803
ω_7	1.0176	37.353	α_{11}	3.3335	20.763	α_{34}	5.4951	24.115

ω_8	1.2216	65.502	α_{12}	3.7174	27.153	α_{35}	5.8102	30.188
ω_9	0.8840	57.053	α_{13}	4.2406	26.082	α_{36}	4.4582	22.725
ω_{11}	1.6268	45.867	α_{14}	4.1455	29.176	α_{37}	4.4086	21.635
ω_{12}	0.6738	36.448	α_{15}	3.8834	25.455	α_{38}	4.3685	6.443
ω_{13}	0.7979	69.871	α_{16}	4.0353	20.874	α_{39}	4.3619	6.7494
ω_{14}	0.8973	33.725	α_{17}	4.2236	29.782	α_{40}	4.1466	18.460
ω_{15}	0.5419	12.041	α_{18}	4.1959	25.305	α_{41}	4.6075	21.011
ω_{16}	0.4540	8.511	α_{19}	3.8100	25.719	α_{42}	4.8625	18.285
ω_{17}	0.4594	18.873	α_{20}	4.4164	21.115	α_{43}	4.3266	24.209
ω_{18}	0.5036	43.634	α_{21}	4.4666	25.637	α_{44}	4.5152	21.315
ω_{19}	0.2299	10.934	α_{22}	4.5327	7.764	λ_2	0.7533	15.156
ω_{20}	0.2986	10.507	α_{23}	4.6204	22.415	λ_3	0.9486	36.105
ω_{21}	0.3489	14.394	α_{24}	4.6215	30.532	β	3.6480	37.870
α_1	4.0813	28.744	α_{25}	4.8048	30.995	δ_1	0.0247	13.049
α_2	4.2889	29.291	α_{26}	4.8227	31.550	δ_2	0.0159	10.197
α_3	4.1733	11.112	α_{27}	5.1552	21.722	δ_3	0.0032	2.429
α_4	4.3660	28.702						

Table A4: Model 2 Overall House Price Index P_2, Land Price Index P_{L2} and Structure Price Index P_{S2}

Quarter	P_2	P_{L2}	P_{S2}	Quarter	P_2	P_{L2}	P_{S2}
2000-1	1.00000	1.00000	1.00000	2005-3	1.04701	1.13207	0.94595
2000-2	1.02266	1.05085	0.98919	2005-4	1.04458	1.13235	0.94054
2000-3	1.00761	1.02254	0.98919	2006-1	1.06559	1.17726	0.93514
2000-4	1.03515	1.06975	0.99459	2006-2	1.07563	1.18166	0.95135
2001-1	1.00551	1.02342	0.98378	2006-3	1.11861	1.26312	0.95135
2001-2	1.00117	1.01052	0.98919	2006-4	1.17166	1.36549	0.95135
2001-3	0.94074	0.90293	0.98378	2007-1	1.17799	1.38299	0.94595
2001-4	0.94753	0.92455	0.97297	2007-2	1.20375	1.43761	0.94054
2002-1	0.98706	1.00559	0.96216	2007-3	1.18259	1.38208	0.95676
2002-2	0.93446	0.90640	0.96757	2007-4	1.23960	1.49321	0.95676
2002-3	0.88163	0.81676	0.96216	2008-1	1.16346	1.35773	0.94054
2002-4	0.93705	0.91084	0.96757	2008-2	1.16367	1.34639	0.95135
2003-1	1.00409	1.03902	0.95676	2008-3	1.21019	1.42361	0.96757
2003-2	0.98870	1.01572	0.95135	2008-4	1.04434	1.09234	0.96216
2003-3	0.95494	0.95151	0.95676	2009-1	1.03267	1.08019	0.95135
2003-4	0.97272	0.98873	0.95135	2009-2	1.03025	1.07036	0.95676
2004-1	0.99752	1.03485	0.95135	2009-3	1.03762	1.06875	0.97297
2004-2	0.99137	1.02807	0.94595	2009-4	1.00210	1.01598	0.95676
2004-3	0.94255	0.93352	0.95135	2010-1	1.06221	1.12892	0.95676
2004-4	1.01822	1.08209	0.94054	2010-2	1.09170	1.19139	0.95135
2005-1	1.02464	1.09440	0.94054	2010-3	1.03058	1.06010	0.96216
2005-2	1.03569	1.11059	0.94595	2010-4	1.05408	1.10631	0.96216

Table A5: Estimated Coefficients for Model 3

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
ω_1	2.1720	46.575	α_7	3.0524	19.437	α_{33}	4.5638	23.738
ω_2	1.0317	39.912	α_8	3.1454	21.707	α_{34}	4.5523	15.929
ω_3	1.2142	38.814	α_9	3.4139	20.278	α_{35}	4.7507	23.788
ω_4	1.0137	11.793	α_{10}	3.1019	21.427	α_{36}	3.5949	18.991
ω_5	0.4055	7.694	α_{11}	2.7167	15.513	α_{37}	3.6080	18.489
ω_6	0.6358	13.024	α_{12}	3.0574	21.675	α_{38}	3.5949	15.800
ω_7	1.0205	41.464	α_{13}	3.5041	19.416	α_{39}	3.5809	18.844
ω_8	1.2354	65.496	α_{14}	3.3939	22.985	α_{40}	3.3792	16.225
ω_9	0.8752	56.853	α_{15}	3.1761	19.769	α_{41}	3.7236	20.300
ω_{11}	1.6534	47.079	α_{16}	3.3300	15.479	α_{42}	3.9279	21.504
ω_{12}	0.6801	39.231	α_{17}	3.4754	23.455	α_{43}	3.4924	12.726
ω_{13}	0.7986	73.161	α_{18}	3.4478	19.539	α_{44}	3.6929	21.283
ω_{14}	0.9205	33.793	α_{19}	3.1505	20.627	λ_2	0.8117	15.869

ω_{15}	0.5377	12.288	α_{20}	3.6162	10.915	λ_3	1.0015	35.792
ω_{16}	0.4473	8.247	α_{21}	3.6453	16.745	κ_2	0.1038	5.421
ω_{17}	0.4555	19.932	α_{22}	3.7602	4.608	κ_3	0.0433	3.318
ω_{18}	0.4964	42.854	α_{23}	3.8321	21.340	κ_4	0.0124	1.555
ω_{19}	0.2149	10.411	α_{24}	3.7670	23.298	β	3.6857	23.881
ω_{20}	0.2895	10.375	α_{25}	3.9370	23.924	δ_1	0.0223	11.218
ω_{21}	0.3409	14.964	α_{26}	3.9708	24.483	δ_2	0.0151	9.312
α_1	3.3850	22.614	α_{27}	4.2376	8.073	δ_3	0.0023	1.701
α_2	3.5200	22.962	α_{28}	4.5457	25.580	ϕ_2	0.0277	0.744
α_3	3.4650	8.242	α_{29}	4.6878	20.300	ϕ_3	−0.0326	-2.227
α_4	3.5976	22.720	α_{30}	4.8662	20.137	ϕ_4	−0.0437	-6.051
α_5	3.4611	22.606	α_{31}	4.6781	6.549			
α_6	3.4286	14.562	α_{32}	5.0071	24.382			

Table A6: Model 3 Overall House Price Index P_3, Land Price Index P_{L3} and Structure Price Index P_{S3}

Quarter	P_3	P_{L3}	P_{S3}	Quarter	P_3	P_{L3}	P_{S3}
2000-1	1.00000	1.00000	1.00000	2005-3	1.04539	1.13209	0.94595
2000-2	1.01625	1.03989	0.98919	2005-4	1.03284	1.11286	0.94054
2000-3	1.00774	1.02363	0.98919	2006-1	1.05613	1.16307	0.93514
2000-4	1.03077	1.06283	0.99459	2006-2	1.06919	1.17308	0.95135
2001-1	1.00456	1.02250	0.98378	2006-3	1.10978	1.25188	0.95135
2001-2	1.00211	1.01288	0.98919	2006-4	1.15580	1.34290	0.95135
2001-3	0.94089	0.90176	0.98378	2007-1	1.17416	1.38487	0.94595
2001-4	0.95050	0.92922	0.97297	2007-2	1.19822	1.43759	0.94054
2002-1	0.98824	1.00854	0.96216	2007-3	1.17809	1.38203	0.95676
2002-2	0.94072	0.91636	0.96757	2007-4	1.22654	1.47922	0.95676
2002-3	0.87576	0.80257	0.96216	2008-1	1.15422	1.34824	0.94054
2002-4	0.93387	0.90323	0.96757	2008-2	1.15850	1.34485	0.95135
2003-1	1.00175	1.03518	0.95676	2008-3	1.19538	1.40348	0.96757
2003-2	0.98149	1.00264	0.95135	2008-4	1.02894	1.06202	0.96216
2003-3	0.94831	0.93830	0.95676	2009-1	1.02513	1.06590	0.95135
2003-4	0.97016	0.98375	0.95135	2009-2	1.02596	1.06203	0.95676
2004-1	0.99290	1.02671	0.95135	2009-3	1.03237	1.05789	0.97297
2004-2	0.98601	1.01855	0.94595	2009-4	0.99389	0.99829	0.95676
2004-3	0.94189	0.93074	0.95135	2010-1	1.04671	1.10005	0.95676
2004-4	1.01010	1.06832	0.94054	2010-2	1.07421	1.16040	0.95135
2005-1	1.01450	1.07691	0.94054	2010-3	1.01655	1.03173	0.96216
2005-2	1.03446	1.11086	0.94595	2010-4	1.04580	1.09096	0.96216

Table A7: Estimated Coefficients for Model 4

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
ω_1	1.8800	39.649	α_9	4.2713	16.589	α_{37}	4.4331	15.561
ω_2	0.9139	39.739	α_{10}	3.8307	16.925	α_{38}	4.4175	16.254
ω_3	1.0587	34.569	α_{11}	3.3846	15.753	α_{39}	4.4676	16.556
ω_4	0.9138	12.722	α_{12}	3.9010	16.996	α_{40}	4.2349	15.959
ω_5	0.3770	9.195	α_{13}	4.3396	16.958	α_{41}	4.6103	16.719
ω_6	0.5766	14.056	α_{14}	4.1868	17.632	α_{42}	4.8746	16.780
ω_7	0.8811	36.306	α_{15}	3.9628	16.601	α_{43}	4.3306	16.387
ω_8	1.1422	72.156	α_{16}	4.0569	16.066	α_{44}	4.5275	16.952
ω_9	0.7884	57.600	α_{17}	4.2484	17.537	λ_2	0.8135	18.683
ω_{11}	1.4181	47.227	α_{18}	4.3038	16.984	λ_3	0.9633	40.256
ω_{12}	0.6770	41.726	α_{19}	3.9044	17.194	κ_2	0.1249	7.904
ω_{13}	0.7923	75.729	α_{20}	4.5513	17.044	κ_3	0.0509	4.060
ω_{14}	0.8150	37.629	α_{21}	4.5148	17.214	κ_4	0.0236	2.977
ω_{15}	0.4971	13.280	α_{22}	4.6399	17.365	τ_2	−0.0035	−1.311
ω_{16}	0.4184	9.476	α_{23}	4.7050	17.148	τ_3	−0.0201	−7.369
ω_{17}	0.4907	23.255	α_{24}	4.6198	17.674	τ_4	−0.0171	−5.496
ω_{18}	0.5856	50.055	α_{25}	4.8524	17.586	μ_2	−0.0008	−0.453

ω_{19}	0.2434	13.220	α_{26}	4.9709	17.656	μ_3	−0.0128	−6.876
ω_{20}	0.2996	12.474	α_{27}	5.1678	17.245	μ_4	−0.0188	−9.065
ω_{21}	0.3641	17.493	α_{28}	5.6178	18.326	β	3.4381	30.267
α_1	4.2002	17.507	α_{29}	5.7830	17.895	δ_1	0.0220	12.322
α_2	4.3367	17.466	α_{30}	6.0430	18.330	δ_2	0.0164	10.763
α_3	4.2781	16.954	α_{31}	5.7929	17.857	δ_3	0.0026	1.971
α_4	4.4558	17.497	α_{32}	6.2191	18.217	ϕ_2	0.0277	1.019
α_5	4.3962	17.572	α_{33}	5.7130	17.616	ϕ_3	−0.0293	−1.919
α_6	4.1720	16.815	α_{34}	5.4707	17.696	ϕ_4	−0.0484	−6.863
α_7	3.7929	16.824	α_{35}	5.7189	18.032			
α_8	3.8968	17.774	α_{36}	4.5691	15.873			

Table A8: Model 4 Overall House Price Index P_4, Land Price Index P_{L4} and Structure Price Index P_{S4}

Quarter	P_4	P_{L4}	P_{S4}	Quarter	P_4	P_{L4}	P_{S4}
2000-1	1.00000	1.00000	1.00000	2005-3	1.04490	1.12020	0.94595
2000-2	1.01372	1.03250	0.98919	2005-4	1.03130	1.09991	0.94054
2000-3	1.00595	1.01856	0.98919	2006-1	1.05927	1.15529	0.93514
2000-4	1.03184	1.06087	0.99459	2006-2	1.08212	1.18349	0.95135
2001-1	1.01915	1.04667	0.98378	2006-3	1.10776	1.23037	0.95135
2001-2	0.99191	0.99329	0.98919	2006-4	1.16545	1.33751	0.95135
2001-3	0.93871	0.90303	0.98378	2007-1	1.18394	1.37685	0.94595
2001-4	0.94809	0.92778	0.97297	2007-2	1.21471	1.43875	0.94054
2002-1	0.99432	1.01693	0.96216	2007-3	1.19022	1.37920	0.95676
2002-2	0.93647	0.91203	0.96757	2007-4	1.24420	1.48067	0.95676
2002-3	0.87228	0.80582	0.96216	2008-1	1.17353	1.36019	0.94054
2002-4	0.94643	0.92876	0.96757	2008-2	1.14956	1.30250	0.95135
2003-1	1.00255	1.03319	0.95676	2008-3	1.18824	1.36160	0.96757
2003-2	0.97921	0.99681	0.95135	2008-4	1.04534	1.08783	0.96216
2003-3	0.95028	0.94349	0.95676	2009-1	1.02307	1.05546	0.95135
2003-4	0.96069	0.96590	0.95135	2009-2	1.02363	1.05173	0.95676
2004-1	0.98643	1.01148	0.95135	2009-3	1.03785	1.06368	0.97297
2004-2	0.99157	1.02466	0.94595	2009-4	1.00020	1.00826	0.95676
2004-3	0.94036	0.92959	0.95135	2010-1	1.04948	1.09764	0.95676
2004-4	1.02247	1.08360	0.94054	2010-2	1.08052	1.16057	0.95135
2005-1	1.01777	1.07492	0.94054	2010-3	1.01775	1.03106	0.96216
2005-2	1.03640	1.10469	0.94595	2010-4	1.04256	1.07794	0.96216

Table A9: Estimated Coefficients for Model 5

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
ω_1	1.9633	39.625	$\alpha_{1,23}$	4.3069	15.778	μ_4	−0.0166	−8.680	$\alpha_{2,23}$	3.5417	5.009
ω_2	0.9057	36.783	$\alpha_{1,24}$	4.1806	16.477	ω_5	0.6785	11.178	$\alpha_{2,24}$	3.7971	5.385
ω_3	1.0575	33.380	$\alpha_{1,25}$	4.3665	16.265	ω_6	1.0016	12.150	$\alpha_{2,25}$	3.8944	12.758
ω_4	0.9048	12.137	$\alpha_{1,26}$	4.4749	16.321	ω_{12}	1.1239	34.884	$\alpha_{2,26}$	4.0142	11.458
ω_7	0.8688	35.194	$\alpha_{1,27}$	4.8225	16.284	ω_{15}	0.8846	15.548	$\alpha_{2,27}$	3.5020	4.894
ω_8	1.1505	72.571	$\alpha_{1,28}$	5.2712	17.114	ω_{16}	0.8193	11.473	$\alpha_{2,28}$	3.9118	13.174
ω_9	0.7772	56.140	$\alpha_{1,29}$	5.3545	16.647	ω_{17}	0.9015	23.072	$\alpha_{2,29}$	4.2516	12.835
ω_{11}	1.4448	47.088	$\alpha_{1,30}$	5.6706	17.103	ω_{19}	0.5338	16.161	$\alpha_{2,30}$	4.2537	6.045
ω_{13}	0.7762	73.276	$\alpha_{1,31}$	5.3258	16.557	ω_{20}	0.6165	14.194	$\alpha_{2,31}$	4.3927	5.774
ω_{14}	0.7968	34.836	$\alpha_{1,32}$	5.8252	17.086	ω_{21}	0.6977	18.855	$\alpha_{2,32}$	4.3479	10.708
$\alpha_{1,1}$	3.6545	15.918	$\alpha_{1,33}$	5.3137	15.935	$\alpha_{2,1}$	3.9412	5.288	$\alpha_{2,33}$	4.1610	4.615
$\alpha_{1,2}$	3.8100	16.064	$\alpha_{1,34}$	5.0382	16.287	$\alpha_{2,2}$	3.9435	10.724	$\alpha_{2,34}$	4.0692	5.695
$\alpha_{1,3}$	3.7829	15.295	$\alpha_{1,35}$	5.4233	16.643	$\alpha_{2,3}$	3.5019	11.765	$\alpha_{2,35}$	3.9107	7.149
$\alpha_{1,4}$	3.9679	16.186	$\alpha_{1,36}$	4.1252	13.967	$\alpha_{2,4}$	3.6617	12.323	$\alpha_{2,36}$	3.5742	10.113
$\alpha_{1,5}$	3.8260	15.959	$\alpha_{1,37}$	4.0739	4.0430	$\alpha_{2,5}$	3.7945	12.647	$\alpha_{2,37}$	3.2497	3.874
$\alpha_{1,6}$	3.7102	15.365	$\alpha_{1,38}$	4.0611	14.675	$\alpha_{2,6}$	3.4356	4.636	$\alpha_{2,38}$	3.3287	9.672
$\alpha_{1,7}$	3.2816	15.089	$\alpha_{1,39}$	4.1395	15.132	$\alpha_{2,7}$	3.3615	5.073	$\alpha_{2,39}$	3.1356	3.393
$\alpha_{1,8}$	3.3929	16.144	$\alpha_{1,40}$	3.8368	14.508	$\alpha_{2,8}$	3.4579	12.098	$\alpha_{2,40}$	3.3051	9.300
$\alpha_{1,9}$	3.8221	15.243	$\alpha_{1,41}$	4.2438	15.356	$\alpha_{2,9}$	3.3557	10.168	$\alpha_{2,41}$	3.3104	4.204
$\alpha_{1,10}$	3.3537	15.457	$\alpha_{1,42}$	4.5134	15.562	$\alpha_{2,10}$	3.5108	10.615	$\alpha_{2,42}$	3.2847	3.482

$\alpha_{1,11}$	2.8895	13.831	$\alpha_{1,43}$	3.8608	14.807	$\alpha_{2,11}$	3.3367	7.438	$\alpha_{2,43}$	3.4985	10.503
$\alpha_{1,12}$	3.4461	15.575	$\alpha_{1,44}$	4.2065	15.509	$\alpha_{2,12}$	3.2431	6.085	$\alpha_{2,44}$	3.2090	9.348
$\alpha_{1,13}$	3.9140	15.793	$\lambda_{1,2}$	0.8949	16.115	$\alpha_{2,13}$	3.3251	9.985	$\lambda_{2,2}$	0.6087	6.293
$\alpha_{1,14}$	3.7744	16.386	$\lambda_{1,3}$	1.0336	32.986	$\alpha_{2,14}$	3.3232	11.164	$\lambda_{2,3}$	0.9214	14.972
$\alpha_{1,15}$	3.5055	15.386	κ_2	0.11591	6.559	$\alpha_{2,15}$	3.7466	9.992	β_1	3.9734	18.808
$\alpha_{1,16}$	3.6281	14.454	κ_3	0.05421	4.017	$\alpha_{2,16}$	3.2874	10.196	δ_1	0.02118	10.485
$\alpha_{1,17}$	3.7883	16.229	κ_4	0.02272	2.667	$\alpha_{2,17}$	3.5232	7.714	δ_2	0.01624	9.106
$\alpha_{1,18}$	3.8840	15.740	τ_2	−0.00373	−1.377	$\alpha_{2,18}$	3.2123	4.691	δ_3	0.00256	1.656
$\alpha_{1,19}$	3.4208	15.792	τ_3	−0.02145	−8.187	$\alpha_{2,19}$	3.7342	10.622	ϕ_2	0.02728	0.749
$\alpha_{1,20}$	4.1079	15.710	τ_4	−0.01531	−5.495	$\alpha_{2,20}$	3.4207	11.054	ϕ_3	−0.02812	−1.880
$\alpha_{1,21}$	4.0751	15.772	μ_2	−0.00096	−0.513	$\alpha_{2,21}$	3.5135	11.298	ϕ_4	−0.05368	−7.104
$\alpha_{1,22}$	4.2874	16.255	μ_3	−0.01421	−7.701	$\alpha_{2,22}$	3.3312	8.016	β_2	2.4777	12.086

Table A10: Model 5 Overall House Price Index P_5, Land Price Index P_{L5} and Land Price Indexes in High and Lower End Wards, $P_{L1,5}$ and $P_{L2,5}$

Quarter	P_5	P_{L5}	$P_{L1,5}$	$P_{L2,5}$	Quarter	P_5	P_{L5}	$P_{L1,5}$	$P_{L2,5}$
2000-1	1.00000	1.00000	1.00000	1.00000	2005-3	1.02137	1.07904	1.17853	0.89864
2000-2	1.01201	1.02999	1.04256	1.00057	2005-4	1.01899	1.07909	1.14395	0.96344
2000-3	0.98713	0.98561	1.03514	0.88854	2006-1	1.03991	1.12072	1.19485	0.98812
2000-4	1.01596	1.03273	1.08577	0.92909	2006-2	1.06383	1.15071	1.22450	1.01852
2001-1	1.00376	1.01939	1.04693	0.96278	2006-3	1.07369	1.16875	1.31962	0.88855
2001-2	0.97668	0.96665	1.01526	0.87170	2006-4	1.13696	1.28512	1.44240	0.99253
2001-3	0.92653	0.88155	0.89797	0.85291	2007-1	1.16185	1.33521	1.46519	1.07875
2001-4	0.93793	0.91004	0.92841	0.87738	2007-2	1.18812	1.38731	1.55170	1.07929
2002-1	0.97445	0.98276	1.04588	0.85145	2007-3	1.17113	1.34302	1.45733	1.11455
2002-2	0.92871	0.89631	0.91770	0.89080	2007-4	1.21379	1.42143	1.59398	1.10318
2002-3	0.86656	0.79076	0.79068	0.84661	2008-1	1.15152	1.32034	1.45402	1.05576
2002-4	0.93102	0.90032	0.94297	0.82288	2008-2	1.12835	1.26750	1.37865	1.03247
2003-1	0.98389	1.00306	1.07101	0.84369	2008-3	1.15670	1.30596	1.48403	0.99227
2003-2	0.96560	0.97492	1.03280	0.84319	2008-4	1.02601	1.06749	1.12881	0.90687
2003-3	0.95058	0.94436	0.95924	0.95061	2009-1	0.99470	1.01921	1.11478	0.82455
2003-4	0.94499	0.93845	0.99277	0.83411	2009-2	1.00168	1.02736	1.11126	0.84458
2004-1	0.97187	0.98656	1.03661	0.89393	2009-3	1.00314	1.01656	1.13273	0.79560
2004-2	0.96727	0.98254	1.06281	0.81506	2009-4	0.97803	0.98446	1.04989	0.83861
2004-3	0.93814	0.92580	0.93607	0.94748	2010-1	1.01763	1.05578	1.16127	0.83994
2004-4	0.99390	1.03393	1.12407	0.86795	2010-2	1.03857	1.09915	1.23503	0.83342
2005-1	0.99595	1.03765	1.11510	0.89147	2010-3	0.99706	1.01227	1.05647	0.88768
2005-2	1.00909	1.05695	1.17320	0.84524	2010-4	1.00904	1.03416	1.17853	0.89864

Table A11: Model 5 Approximate Stock and Sales House Price Indexes, P_{K5} and P_5, and Approximate Stock and Sales Land Price Indexes, P_{KL5} and P_{L5}.

Quarter	P_{K5}	P_5	P_{KL5}	P_{L5}	Quarter	P_{K5}	P_5	P_{KL5}	P_{L5}
2000-1	1.00000	1.00000	1.00000	1.00000	2005-3	1.02083	1.02137	1.08079	1.07904
2000-2	1.01068	1.01201	1.02790	1.02999	2005-4	1.01849	1.01899	1.08092	1.07909
2000-3	0.98628	0.98713	0.98395	0.98561	2006-1	1.03927	1.03991	1.12266	1.12072
2000-4	1.01484	1.01596	1.03106	1.03273	2006-2	1.06309	1.06383	1.15257	1.15071
2001-1	1.00253	1.00376	1.01754	1.01939	2006-3	1.07226	1.07369	1.16909	1.16875
2001-2	0.97583	0.97668	0.96513	0.96665	2006-4	1.13680	1.13696	1.28531	1.28512
2001-3	0.92740	0.92653	0.88224	0.88155	2007-1	1.15935	1.16185	1.33025	1.33521
2001-4	0.93833	0.93793	0.91059	0.91004	2007-2	1.18831	1.18812	1.38673	1.38731
2002-1	0.97095	0.97445	0.97798	0.98276	2007-3	1.16826	1.17113	1.33763	1.34302
2002-2	0.93466	0.92871	0.90831	0.89631	2007-4	1.21544	1.21379	1.42260	1.42143
2002-3	0.87779	0.86656	0.81021	0.79076	2008-1	1.14845	1.15152	1.31495	1.32034
2002-4	0.93062	0.93102	0.90104	0.90032	2008-2	1.12150	1.12835	1.25776	1.26750
2003-1	0.97612	0.98389	0.99163	1.00306	2008-3	1.15900	1.15670	1.31230	1.30596
2003-2	0.95981	0.96560	0.96659	0.97492	2008-4	1.01166	1.02601	1.05131	1.06749
2003-3	0.95646	0.95058	0.95623	0.94436	2009-1	0.98582	0.99470	1.01343	1.01921
2003-4	0.94359	0.94499	0.93737	0.93845	2009-2	0.99084	1.00168	1.01814	1.02736
2004-1	0.97103	0.97187	0.98678	0.98656	2009-3	0.99631	1.00314	1.01500	1.01656
2004-2	0.96280	0.96727	0.97629	0.98254	2009-4	0.96751	0.97803	0.97611	0.98446

2004-3	**0.94508**	**0.93814**	**0.94005**	**0.92580**	**2010-1**	**1.00801**	**1.01763**	**1.04906**	**1.05578**
2004-4	**0.99279**	**0.99390**	**1.03463**	**1.03393**	**2010-2**	**1.03100**	**1.03857**	**1.09479**	**1.09915**
2005-1	**0.99411**	**0.99595**	**1.03701**	**1.03765**	**2010-3**	**0.98180**	**0.99706**	**0.99753**	**1.01227**
2005-2	**1.00854**	**1.00909**	**1.05867**	**1.05695**	**2010-4**	**1.00174**	**1.00904**	**1.03343**	**1.03416**

Table A12: Rolling Window Overall House Price Index P_{RW}, Land Price Index P_{LRW} and Land Price Indexes in High and Lower End Wards, $P_{L1,RW}$ and $P_{L2,RW}$

Quarter	P_{RW}	P_{LRW}	$P_{L1,RW}$	$P_{L2,RW}$	Quarter	P_{RW}	P_{LRW}	$P_{L1,RW}$	$P_{L2,RW}$
2000-1	1.00000	1.00000	1.00000	1.00000	2005-3	1.01411	1.05945	1.14988	0.89238
2000-2	1.01228	1.02811	1.03734	1.00583	2005-4	1.02054	1.07401	1.13421	0.96513
2000-3	0.98593	0.98375	1.03037	0.89096	2006-1	1.03500	1.10267	1.16709	0.98580
2000-4	1.01432	1.02778	1.08182	0.92161	2006-2	1.05539	1.12605	1.19204	1.00633
2001-1	1.00730	1.02340	1.05642	0.95697	2006-3	1.06070	1.13554	1.27315	0.86857
2001-2	0.97594	0.96672	1.01630	0.86941	2006-4	1.12514	1.25442	1.39276	0.98600
2001-3	0.93666	0.90428	0.93207	0.85243	2007-1	1.15159	1.30827	1.42802	1.06526
2001-4	0.94056	0.91810	0.94208	0.87492	2007-2	1.18679	1.37925	1.54072	1.07092
2002-1	0.97623	0.98487	1.04244	0.86578	2007-3	1.16342	1.32029	1.42645	1.10303
2002-2	0.92730	0.89828	0.92117	0.88636	2007-4	1.21164	1.41317	1.57841	1.10117
2002-3	0.87064	0.80737	0.81015	0.85516	2008-1	1.15232	1.31305	1.44207	1.05369
2002-4	0.93344	0.90860	0.95137	0.82852	2008-2	1.12192	1.24189	1.35716	1.00448
2003-1	0.97572	0.98697	1.05123	0.83457	2008-3	1.15304	1.28652	1.45655	0.97959
2003-2	0.96415	0.97133	1.02588	0.84675	2008-4	1.02186	1.03323	1.09920	0.87706
2003-3	0.94577	0.93724	0.95491	0.93340	2009-1	0.99551	0.99304	1.10392	0.78447
2003-4	0.93704	0.92610	0.97653	0.82754	2009-2	0.99932	0.99525	1.09040	0.80474
2004-1	0.97739	0.99425	1.04734	0.89105	2009-3	0.99581	0.97454	1.09572	0.75124
2004-2	0.96862	0.98314	1.06454	0.80906	2009-4	0.96849	0.93868	1.00325	0.80533
2004-3	0.93445	0.92147	0.93313	0.93794	2010-1	1.00796	1.01008	1.12593	0.78503
2004-4	0.98957	1.02155	1.11045	0.85348	2010-2	1.03134	1.05965	1.19931	0.79429
2005-1	1.00220	1.04308	1.12860	0.87982	2010-3	0.98908	0.96868	1.02299	0.83814
2005-2	1.00732	1.04794	1.16088	0.83817	2010-4	0.99605	0.98170	1.10245	0.76350

Table A13: Estimated Coefficients for Model 6

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
β	0.44108	40.32	α_{22}	-0.03206	-1.557	ω_1	2.5971	100.4
γ	0.49710	56.01	α_{23}	-0.01547	-0.744	ω_2	2.1231	98.70
δ	-0.09662	-27.70	α_{24}	0.00220	0.110	ω_3	2.2229	89.74
α_2	0.01980	1.041	α_{25}	0.01510	0.782	ω_4	2.0598	43.58
α_3	-0.00977	-0.476	α_{26}	0.03062	1.565	ω_5	1.6997	49.55
α_4	0.02905	1.476	α_{27}	0.05679	2.699	ω_6	1.8396	56.91
α_5	0.00439	0.233	α_{28}	0.07446	4.000	ω_7	2.1286	99.60
α_6	-0.03044	-1.498	α_{29}	0.10867	5.589	ω_8	2.2722	124.6
α_7	-0.03968	-1.975	α_{30}	0.11894	5.744	ω_9	2.0451	114.4
α_8	-0.06480	-3.368	α_{31}	0.11589	5.610	ω_{10}	2.1472	131.8
α_9	-0.03407	-1.585	α_{32}	0.15764	7.989	ω_{11}	2.4110	104.6
α_{10}	-0.06956	-3.558	α_{33}	0.09925	4.974	ω_{12}	1.9484	103.8
α_{11}	-0.08096	-3.821	α_{34}	0.09112	4.370	ω_{13}	2.0261	119.6
α_{12}	-0.06539	-3.447	α_{35}	0.09137	4.407	ω_{14}	2.0362	92.80
α_{13}	-0.04697	-2.261	α_{36}	-0.00664	-0.297	ω_{15}	1.8462	65.06
α_{14}	-0.05446	-2.815	α_{37}	-0.03158	-1.400	ω_{16}	1.7700	52.71
α_{15}	-0.05867	-2.686	α_{38}	-0.00740	-0.345	ω_{17}	1.7599	88.92
α_{16}	-0.05199	-2.398	α_{39}	-0.04020	-1.921	ω_{18}	1.7685	105.1
α_{17}	-0.04769	-2.376	α_{40}	-0.03208	-1.448	ω_{19}	1.4993	79.25
α_{18}	-0.04756	-2.406	α_{41}	-0.01766	-0.858	ω_{20}	1.6137	75.76
α_{19}	-0.06290	-3.153	α_{42}	-0.01507	-0.746	ω_{21}	1.6584	82.44
α_{20}	-0.03342	-1.640	α_{43}	-0.01534	-0.730			
α_{21}	-0.01566	-0.823	α_{44}	-0.02218	-1.110			

Table A14: Model 6 and 7 House Price Indexes for Tokyo

Quarter	P_6	P_7	Quarter	P_6	P_7	Quarter	P_6	P_7
2000-1	1.00000	1.00000	2003-4	0.94934	0.93746	2007-3	1.12287	1.12422
2000-2	1.02001	1.01284	2004-1	0.95342	0.93431	2007-4	1.17074	1.15983
2000-3	0.99027	0.98416	2004-2	0.95355	0.94588	2008-1	1.10435	1.09802
2000-4	1.02948	1.01977	2004-3	0.93903	0.92946	2008-2	1.09541	1.08060
2001-1	1.00440	1.00459	2004-4	0.96713	0.96107	2008-3	1.09568	1.08083
2001-2	0.97002	0.96648	2005-1	0.98446	0.98389	2008-4	0.99338	0.99594
2001-3	0.96110	0.95013	2005-2	0.96845	0.96045	2009-1	0.96891	0.95382
2001-4	0.93725	0.93426	2005-3	0.98464	0.98378	2009-2	0.99262	0.97806
2002-1	0.96650	0.95806	2005-4	1.00220	0.98825	2009-3	0.96059	0.95341
2002-2	0.93280	0.92990	2006-1	1.01522	1.00744	2009-4	0.96843	0.95632
2002-3	0.92223	0.90646	2006-2	1.03110	1.02775	2010-1	0.98249	0.96829
2002-4	0.93670	0.93545	2006-3	1.05844	1.05257	2010-2	0.98504	0.97775
2003-1	0.95411	0.95196	2006-4	1.07731	1.06879	2010-3	0.98477	0.97289
2003-2	0.94700	0.92611	2007-1	1.11479	1.11544	2010-4	0.97806	0.96799
2003-3	0.94302	0.93955	2007-2	1.12630	1.13143			

Table A15: Estimated Coefficients for Model 7

Name	Est Coef	T Stat	Name	Est Coef	T Stat	Name	Est Coef	T Stat
β	0.42882	38.65	α_{19}	-0.07315	-3.907	α_{43}	-0.02749	-1.392
γ	0.52920	62.73	α_{20}	-0.03970	-2.075	α_{44}	-0.03253	-1.732
δ	-0.08885	-26.50	α_{21}	-0.01624	-0.909	ω_1	2.7576	92.38
κ	0.10277	11.42	α_{22}	-0.04035	-2.089	ω_2	2.3117	83.98
ϕ	-0.00190	-7.43	α_{23}	-0.01635	-0.838	ω_3	2.3799	81.23
τ	-0.00106	-20.64	α_{24}	-0.01182	-0.631	ω_4	2.1596	45.16
μ	-0.00007	-16.87	α_{25}	0.07416	0.409	ω_5	1.8569	49.69
α_2	0.01276	0.715	α_{26}	0.02737	1.490	ω_6	2.0014	56.11
α_3	-0.01596	-0.828	α_{27}	0.05123	2.592	ω_7	2.2714	83.22
α_4	0.01957	1.060	α_{28}	0.06653	3.803	ω_8	2.4986	94.55
α_5	0.00458	0.259	α_{29}	0.10925	5.988	ω_9	2.2402	84.61
α_6	-0.03409	-1.788	α_{30}	0.12348	6.351	ω_{10}	2.4074	90.85
α_7	-0.05116	-2.714	α_{31}	0.11709	6.038	ω_{11}	2.5953	90.93
α_8	-0.06800	-3.766	α_{32}	0.01483	8.001	ω_{12}	2.1895	81.14
α_9	-0.04284	-2.124	α_{33}	0.09351	4.989	ω_{13}	2.2736	85.88
α_{10}	-0.07268	-3.961	α_{34}	0.07752	3.960	ω_{14}	2.2275	79.26
α_{11}	-0.09821	-4.936	α_{35}	0.07773	3.994	ω_{15}	2.0107	61.46
α_{12}	-0.06673	-3.749	α_{36}	-0.00407	-0.194	ω_{16}	1.9037	52.32
α_{13}	-0.04923	-2.526	α_{37}	-0.04728	-2.230	ω_{17}	2.0317	68.52
α_{14}	-0.07676	-4.223	α_{38}	-0.02219	-1.099	ω_{18}	2.0899	72.81
α_{15}	-0.06236	-3.042	α_{39}	-0.04771	-2.427	ω_{19}	1.7414	62.21
α_{16}	-0.06458	-3.172	α_{40}	-0.04466	-2.143	ω_{20}	1.8376	63.23
α_{17}	-0.06794	-3.604	α_{41}	-0.03222	-1.664	ω_{21}	1.9016	65.44
α_{18}	-0.05564	-2.998	α_{42}	-0.02250	-1.184			

References

Bostic, R.W., S.D. Longhofer and C.L. Readfearn (2007), “Land Leverage: Decomposing Home Price Dynamics”, *Real Estate Economics* 35:2, 183-2008.

Clapp, J.M. (1980), “The Elasticity of Substitution for Land: The Effects of Measurement Errors”, *Journal of Urban Economics* 8, 255-263.

Court, A. T. (1939), “Hedonic Price Indexes with Automotive Examples”, pp. 98-117 in *The Dynamics of Automobile Demand*, New York: General Motors Corporation.

Davis, M.A. and J. Heathcote (2007), "The Price and Quantity of Residential Land in the United States", *Journal of Monetary Economics* 54. 2595-2620.

de Haan, J. (2003), "Time Dummy Approaches to Hedonic Price Measurement", Paper presented at the Seventh Meeting of the International Working Group on Price Indices, (Ottawa Group), May 27-29, 2003, INSEE, Paris.

de Haan, J. (2008), "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches", Centre for Applied Economic Research Working Paper 2008/01, Faculty of Economics and Commerce, University of New South Wales, Sydney, Australia.

de Haan, J. (2009), "Comment on Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 196-200 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees and C.R. Hulten (eds.), Studies in Income and Wealth 70, Chicago: University of Chicago Press.

de Haan, J., W.E. Diewert (eds.) (2011), *Handbook on Residential Property Price Indexes*, Luxembourg: Eurostat.
http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/hps/rppi_handbook

de Haan, J. and H. van der Grient (2011), "Eliminating Chain Drift in Price Indexes based on Scanner Data", *Journal of Econometrics* 161, 36-46.

de Haan, J. and F. Krsinich (2012), "Scanner Data and the Treatment of Quality Change in Rolling Year GEKS Price Indexes", paper presented at the EMG Workshop 2012, November 23, 2012, Australian School of Business, University of New South Wales, Crowne Plaza Hotel, Coogee Beach, Sydney.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited", *Journal of Productivity Analysis* 3, 211-248.

Diewert, W.E. (2003a), "Hedonic Regressions: A Consumer Theory Approach", pp. 317-348 in *Scanner Data and Price Indexes*, R.C. Feenstra and M.D. Shapiro (eds.), Studies in Income and Wealth 64, Chicago: University of Chicago.

Diewert, W.E. (2003b), "Hedonic Regressions: A Review of Some Unresolved Issues", paper presented at the 7th Meeting of the Ottawa Group, Paris, May 27-29.

Diewert, W.E. (2010), "Alternative Approaches to Measuring House Price Inflation", Discussion Paper 10-10, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E. (2011), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions", pp. 87-116 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Diewert, W.E., J. de Haan and R. Hendriks (2011a), "The Decomposition of a House Price Index into Land and Structures Components: A Hedonic Regression Approach", *The Valuation Journal* 6, (2011), 58-106.

Diewert, W.E., J. de Haan and R. Hendriks (2011b), "Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components"' Discussion Paper 11-01, Department of Economics, University of British Columbia, Vancouver, Canada, V6T1Z1. Forthcoming in *Econometric Reviews*.

Diewert, W.E., S. Heravi and M. Silver (2009), "Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 161-196 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees and C.R. Hulten (eds.), Studies in Income and Wealth 70, Chicago: University of Chicago Press.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Francke, M.K. (2008), "The Hierarchical Trend Model", pp. 164-180 in *Mass Appraisal Methods: An International Perspective for Property Valuers*, T. Kauko and M. Damato (eds.), Oxford: Wiley-Blackwell.

Francke, M.K. and G.A. Vos (2004), "The Hierarchical Trend Model for Property Valuation and Local Price Indices", *Journal of Real Estate Finance and Economics* 28:2/3, 179-208.

Gyourko, J. and A. Saiz (2004), "Reinvestment in the Housing Stock: The Role of Construction Costs and the Supply Side", *Journal of Urban Economics* 55, 238-256.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

Hill, P. (ed.) (2004), *Consumer Price Index Manual: Theory and Practice*, Geneva: International Labour Office.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Koev, E. and J.M.C. Santos Silva (2008), "Hedonic Methods for Decomposing House Price Indices into Land and Structure Components", unpublished paper, Department of Economics, University of Essex, England, October.

Lowe, J. (1823), *The Present State of England in Regard to Agriculture, Trade and Finance*, second edition, London: Longman, Hurst, Rees, Orme and Brown.

McMillen, D.P. (2003), "The Return of Centralization to Chicago: Using Repeat Sales to Identify Changes in House Price Distance Gradients", *Regional Science and Urban Economics* 33, 287-304.

Muth, R.F. (1971), "The Derived Demand for Urban Residential Land", *Urban Studies* 8, 243-254.

Rambaldi, A.N., R.R.J McAllister, K. Collins and C.S. Fletcher (2010), "Separating Land from Structure in Property Prices: A Case Study from Brisbane Australia", School of Economics, The University of Queensland, St. Lucia, Queensland 4072, Australia.

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

Schwann, G.M. (1998), "A Real Estate Price Index for Thin Markets", *Journal of Real Estate Finance and Economics* 16:3, 269-287.

Shimizu, C. and K.G. Nishimura (2007), "Pricing Structure in Tokyo Metropolitan Land Markets and Its Structural Changes: Pre-bubble, Bubble, and Post-bubble Periods," *Journal of Real Estate Finance and Economics* 35(4), 495-496.

Shimizu, C., K.G. Nishimura and T. Watanabe (2010), "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures", *Journal of Economics and Statistics* 230/6, 792-813.

Shimizu, C., K.G. Nishimura and T. Watanabe (2012), "House Prices from Magazines, Realtors, and the Land Registry", *Property Market and Financial Stability*, BIS Papers No.64, Bank of International Settlements, March 2012, 29-38.

Shimizu, C., H. Takatsuji, H. Ono and Nishimura (2010), "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices", *International Journal of Housing Markets and Analysis* 3:4, 351-368.

Statistics Portugal (Instituto Nacional de Estatistica) (2009), "Owner-Occupied Housing: Econometric Study and Model to Estimate Land Prices, Final Report", paper presented to the Eurostat Working Group on the Harmonization of Consumer Price Indices", March 26-27, Luxembourg: Eurostat.

Thorsnes, P. (1997), "Consistent Estimates of the Elasticity of Substitution between Land and Non-Land Inputs in the Production of Housing", *Journal of Urban Economics* 42, 98-108.

White, K.J. (2004), *Shazam: User's Reference Manual, Version 10*, Vancouver, Canada: Northwest Econometrics Ltd.