Residential Property Price Indexes for Tokyo

Erwin Diewert
University of British Columbia and
University of New South Wales
and
Chihiro Shimizu
National University of Singapore and
University of British Columbia

Oct 16, 2015
(The Canon Institute for Global Studies)

Introduction

- **For many purposes, it is useful to have separate price (and quantity) indexes for residential housing; e.g., this information is required in order to construct SNA (real) balance sheets.**
- **The Eurostat *Residential Property Price Indices Handbook* suggested a hedonic regression model (the builder's model) that could be used to decompose the transaction prices of residential houses into separate land and structure components and the method was tested using real estate sales data for the small town of "A" in the Netherlands.**
- **However, an open question is: can the builder's model be applied to a large urban center where neighbourhood effects are going to be much more important than they were for the town of "A"?**

Introduction (cont)

- **To answer this question, we attempted to implement the builder's model for the large city of Tokyo.**
- **Our data on sales of residential houses in Tokyo was not dense enough for us to implement the basic builder's model for individual neighbourhoods (or Wards) of Tokyo so in the present paper, we used Ward dummy variables to take into account neighbourhood effects on the price of land.**
- **In the *RPPI Handbook*, it was found that information on the sales price of a house in the town of "A" along with information on the lot area, the structure area and the age of the structure was sufficient to explain about 85% of the variation in house prices. (These are the basic characteristics).**
- **Another innovation in the present paper is the extension of the basic *RPPI* model to include other house characteristics. These extra characteristics were added as spline variables in order to achieve a better description of the data.**

The Variables

- **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 Data

- **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.**

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

The Data (cont)

- **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: $2 \leq V \leq 20$; $0.5 \leq S \leq 2.5$; $0.5 \leq L \leq 2.5$; $1 \leq A \leq 50$; ; $2 \leq NB \leq 8$; $2.5 \leq W \leq 9$.**
- **In order to eliminate the multicollinearity problem between the lot size L and floor space area S for an individual house, we assumed that the value of a new structure in any quarter was proportional to a Construction Cost Price Index for Tokyo.**

The Data (cont)

- **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.**

The Basic Builder's Model

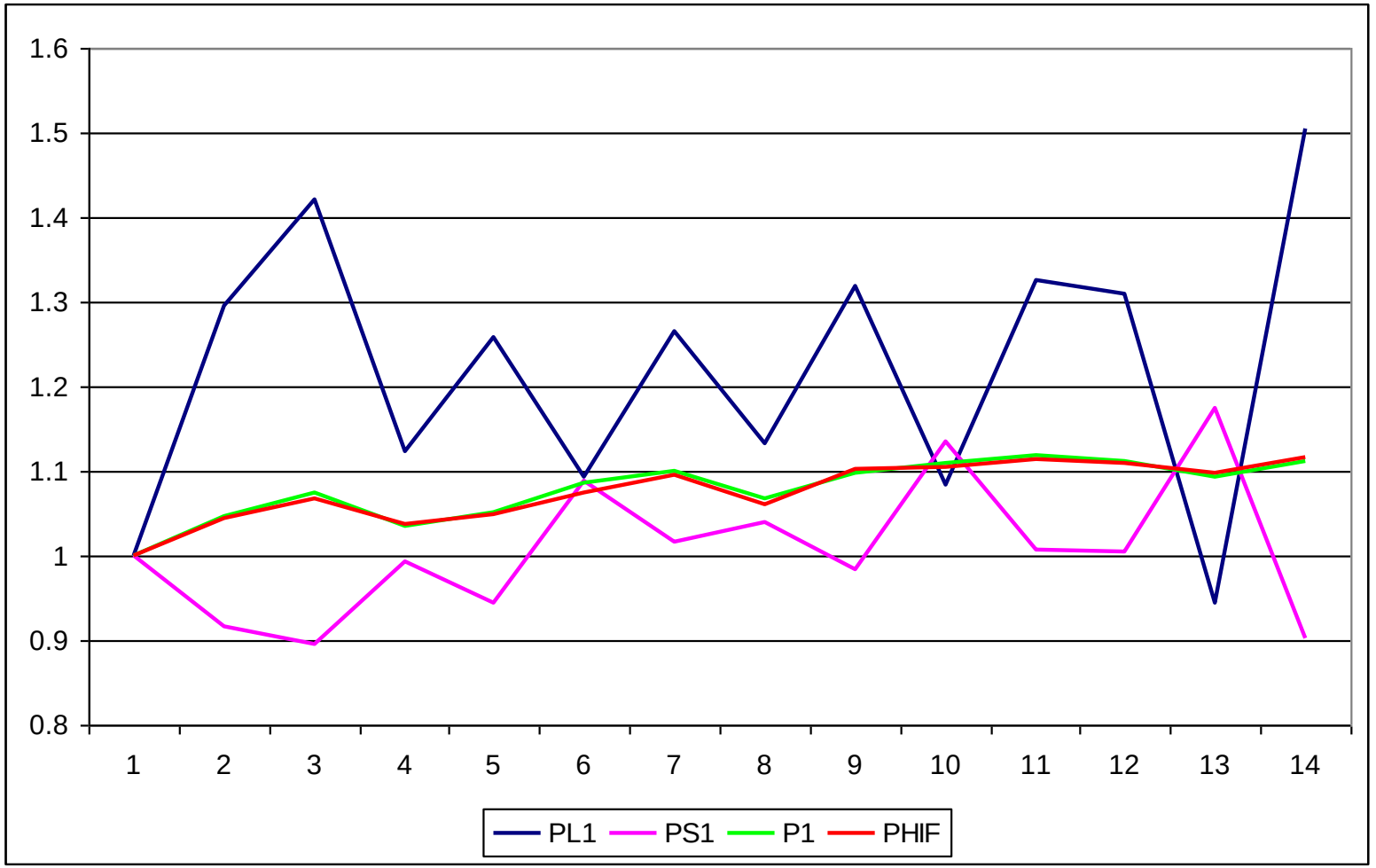
Consider the following hedonic regression model for quarter t:

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

where the parameter δ_t reflects the *net depreciation rate* as the structure ages one additional year, the parameter α_t is the price of land in Tokyo in quarter t and β_t is the price of a new structure in period t. Note that we are assuming straight line depreciation here. To eliminate the multicollinearity problem, we combined all of the 44 quarterly regressions into a single regression and also used our exogenous price of new construction, p_{Ct} (Note that the β_t have been replaced by βp_{Ct}):

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

The Multicollinearity Problem Illustrated: Figure 9.1 from the RPPI Handbook: The Price of Land P_{L1}, the Price of Structures P_{S1}, the Overall Fisher House Price Index P_1 and the Fisher Hedonic Imputation House Price Index P_{HIF}



The Basic Builder's Model: Results

- **For the model defined by equations (3), 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.**
- **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.**
- **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 so we would expect substantial variations in the price of land across the various neighborhoods.**

The Basic Builder's Model with Ward Dummy Variables

- **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).**
- **We now modify the model defined by (3) to allow the *level* of land prices to differ across the 21 Wards of Tokyo:**

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

- **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). We call the model defined by (5) and (6) Model 1.**
- **However, not all parameters can be identified so we impose the following normalization on the parameters:**

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

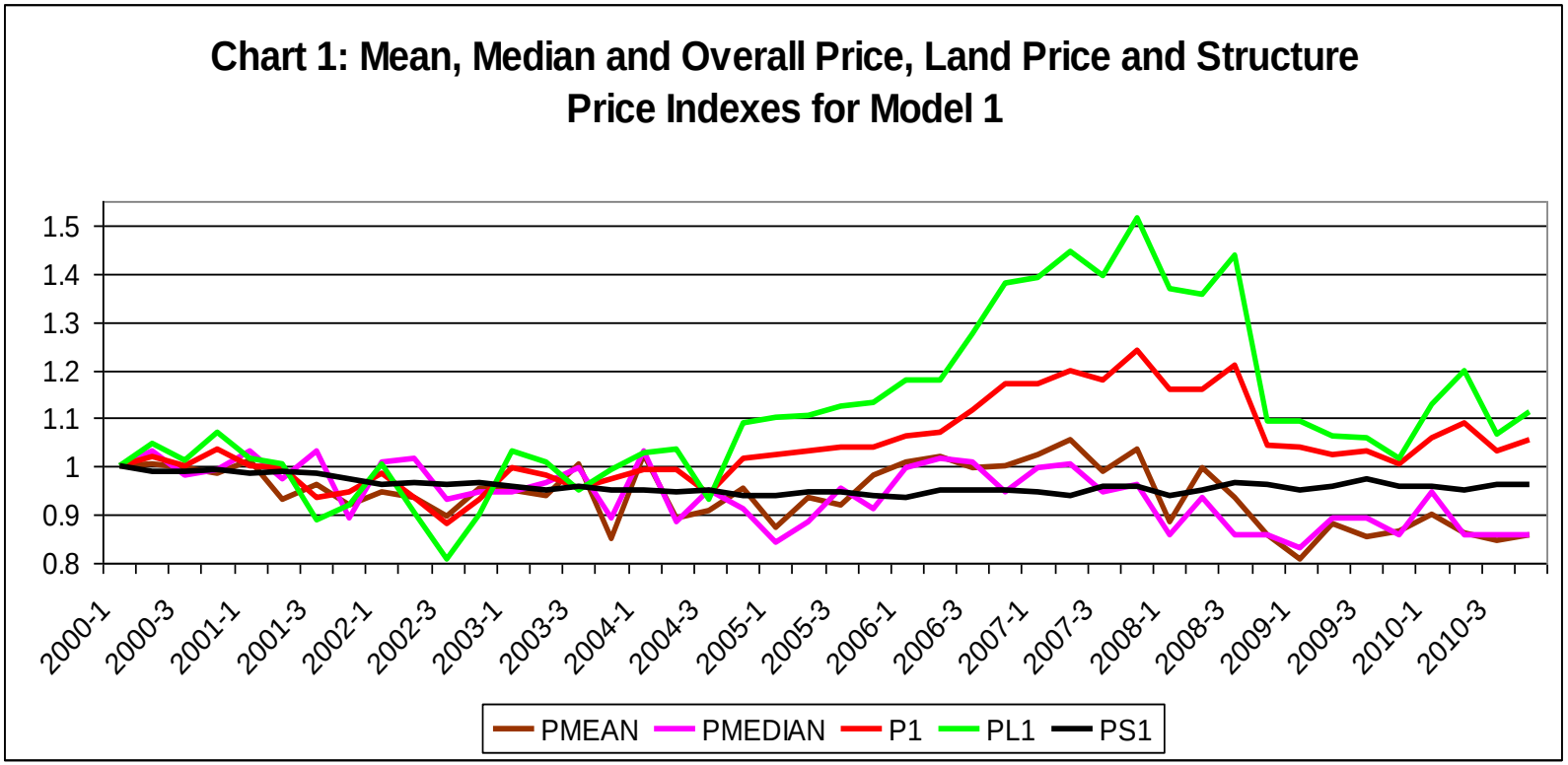
- **The tenth ward, Setagaya, has the most transactions in our sample (1158 transactions over the sample period).**

The Basic Builder's Model with Ward Dummy Variables: Results

- **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.**
- **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.**
- **However, we also required an exogenous house construction price index to implement our method.**
- **We will omit the details on how the results of the hedonic regression (5) were used to construct separate land and structure price indexes. Chained Fisher indexes were used to aggregate the land and structure components into an overall house price index. The following Chart plots these indexes.**

The Basic Builder's Model with Ward Dummy Variables: Results

- **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 are graphed in Chart 1.**
- **We have also computed the quarterly mean and median house prices transacted in each quarter.**



Splines on Lot Size and on the Age of the Structure

- **In this model, we allow the cost of land to be a *piecewise linear function* of the area of the land that the structure sits on.**
- **We also allow the net depreciation of the structure to be a *piecewise linear function* of the age of the structure.**
- **We first look at the lot sizes of the houses in our sample and divide up the observations into roughly 3 equally sized groups, depending on the lot size.**
- **Recall that we have restricted the range of the land variable to $0.5 \leq L_{tn} \leq 2.5$. 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 $0.5 \leq L_{tn} < 0.77$,

1833 observations fell into the interval $0.77 \leq L_{tn} < 1.10$ and

1884 observations fell into the interval $1.1 \leq L_{tn} \leq 2.5$

Splines on Lot Size and on the Age of the Structure (cont)

- 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:

(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}, $f_L(L_{tn})$:

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

where the λ_k are 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.

Splines on Lot Size and on the Age of the Structure (cont)

- **Now divide up our 5578 observations into 3 roughly equal groups based on the age of the structure.**
- **We chose the house ages where there is a change in the marginal depreciation rate to be $A_1 \equiv 10$ and $A_2 \equiv 20$. We found that 2085 observations fell into the interval $0 \leq A_{tn} < 10$, 1996 observations fell into the interval $10 \leq A_{tn} < 20$ and 1497 observations fell into the interval $20 \leq A_{tn} \leq 50$.**
- **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:**

(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.

Splines on Lot Size and on the Age of the Structure (cont)

- 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:

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

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.
- Our **new Model 2 hedonic regression model** is defined as follows: For t = 1,...,44 and n = 1,...,N(t):

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

where the functions f_L and g_A are defined above by (10) and (12) and ε_{tn} is an error term. We also require the normalizations:

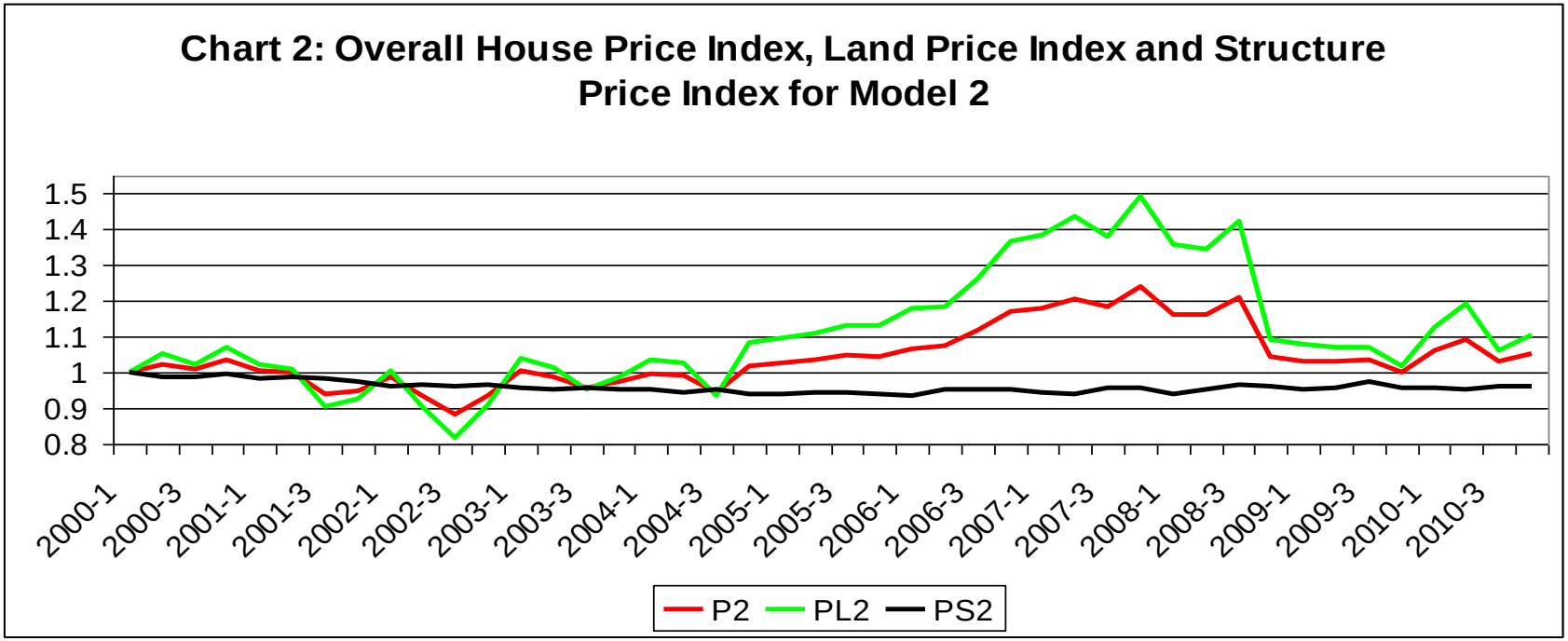
$$(14)\ \omega_{10} = 1;\ \lambda_1 = 1.$$

Splines on Lot Size and on the Age of the Structure: Results

- **There are 44+1+20+2+3 = 70 unknown parameters to be estimated (5 additional parameters over our previous model).**
- **The nonlinear regression model defined by (11) and (12) is our *Model 2*.**
- **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.**
- **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. Thus the price of land per unit lot area is highest for small lots.**
- **Our estimated net depreciation rate parameters for Model 2 were $\delta_1 = 0.0247$, $\delta_2 = 0.0159$ and $\delta_3 = 0.0032$. Thus for houses less than 10 years old, the annual net depreciation rate is 2.47%, for houses between 10 and 20, the marginal depreciation rate drops to 1.59% and for old houses, the marginal rate drops even lower to 0.32% per year.**

Splines on Lot Size and on the Age of the Structure: Results

- **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. There is little difference from the indexes in Chart 1.**



Quality Adjustment for the Number of Bedrooms and Lot Width

- **In this Model 3, we will add some additional explanatory variables to our regression: the width of the lot and the number of bedrooms in the structure.**
- **We will again use piecewise linear functions of the width and bedroom variables to describe how the property price varies as the amounts of these characteristics vary.**
- **Omitting the details of how the spline functions were defined, our Model 3 regression is the following one:**

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

with the following normalizations on the parameters:

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

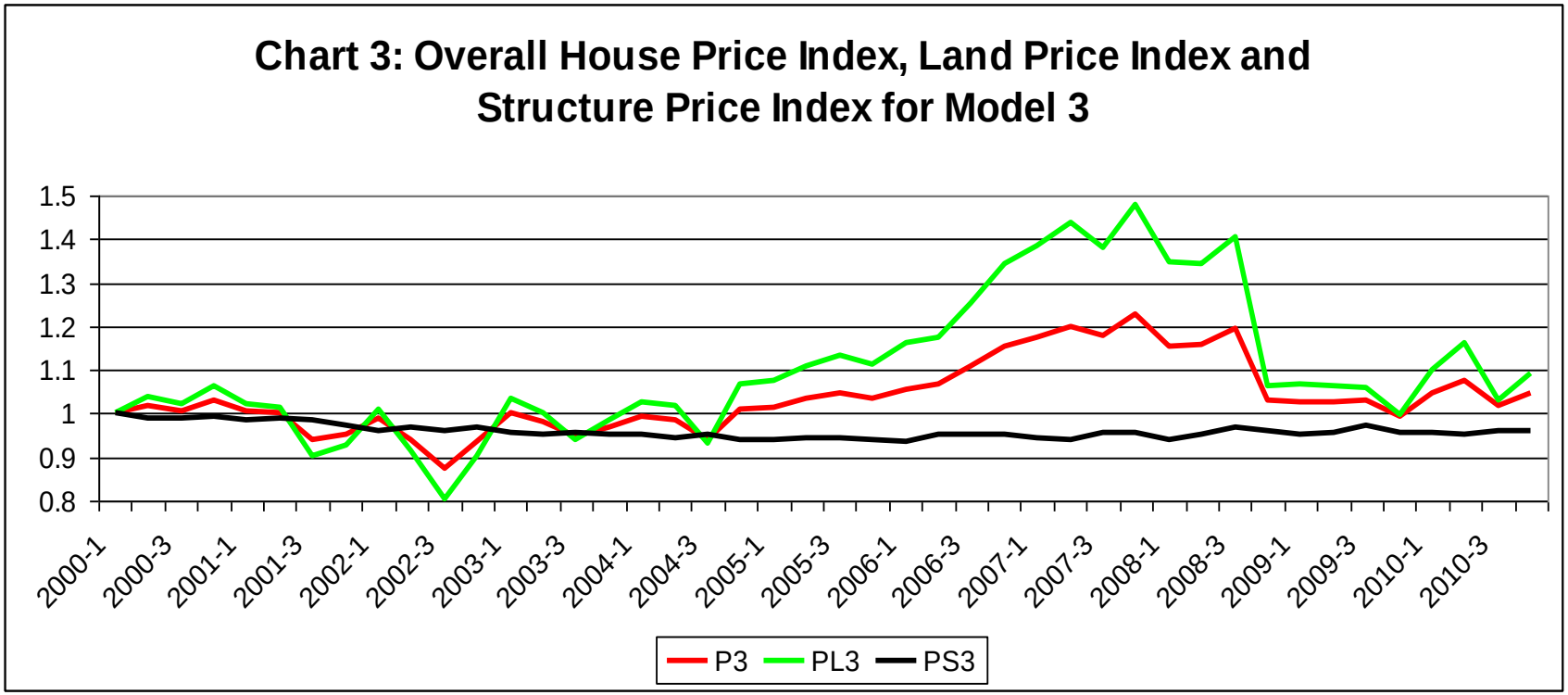
- **We add 6 parameters to Model 2 for a total of 76 parameters to estimate for Model 3. The Frontage variable F_{tn} affects the price of land while B_{tn} affects the price of structures.**

Quality Adjustment for the Number of Bedrooms and Lot Width: Results

- **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.**
- **Thus adding the 3 extra lot width parameters and the 3 extra bedroom parameters is well justified in terms of improving the descriptive power of the model.**
- **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.**

Quality Adjustment for the Number of Bedrooms and Lot Width: Results

- **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.**



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

- **Recall that the sample range of TW (walking time to the nearest subway station) was 2 to 29 minutes while the sample range of TT (subway travel time to the Central Tokyo Station) was 4 to 48 minutes.**
- **Define the following transformations of these variables:**

(27) $M_{tn} \equiv TW_{tn} - 2$;

(30) $T_{tn} \equiv TT_{tn} - 4$.

- **As usual, the observations were broken up into 3 groups for the walking time variable and 3 groups for the subway travel time variable, dummy variables were defined, leading to the following Model 4:**

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

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

Quality Adjustment for the Nearness to Subway Lines and Subway Travel Time: Results

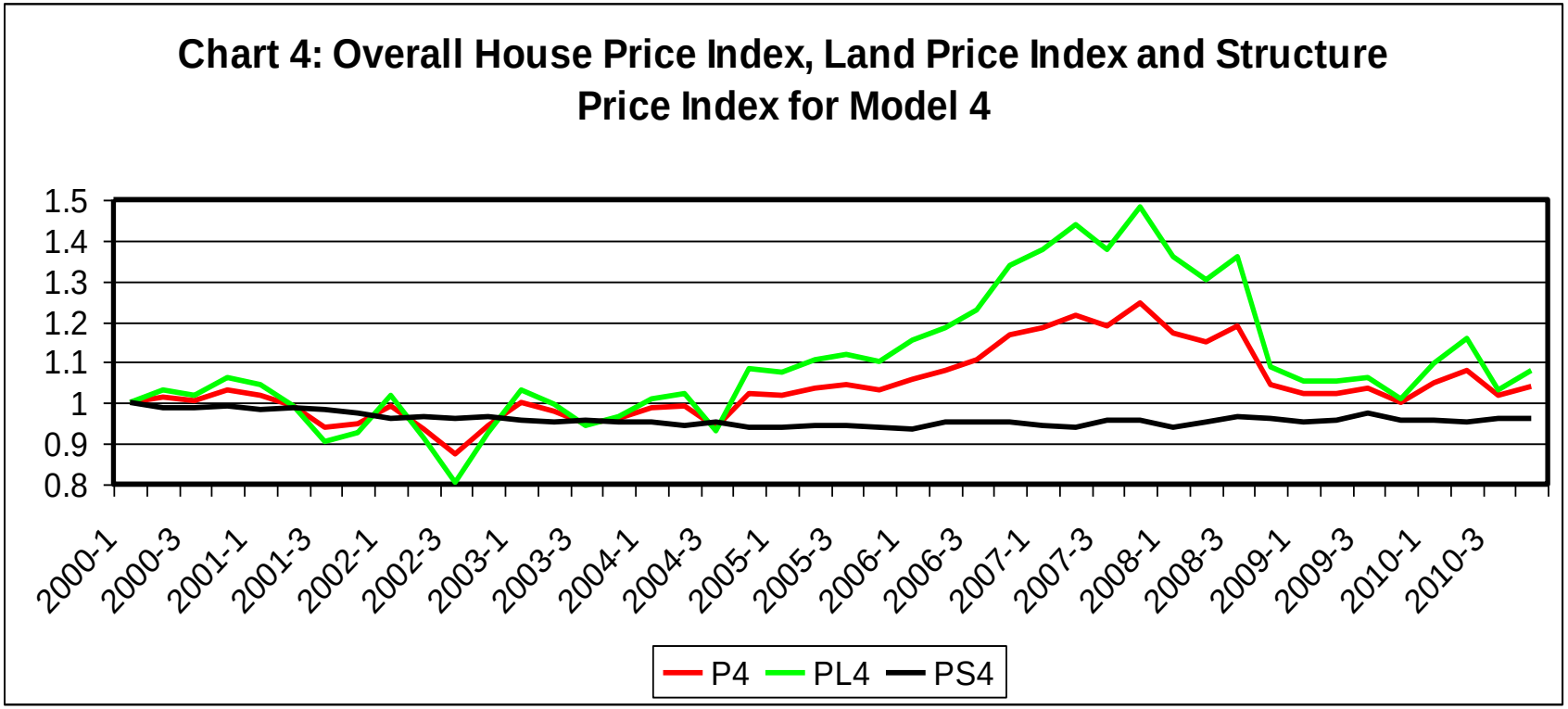
- **The R^2 for Model 4 turned out to be 0.8417 and the log likelihood was –8815.9, a very large increase of 269.4 over the Model 3 log likelihood.**
- **Thus adding the 3 extra walking time parameters and the 3 extra travel time to Tokyo station parameters provides a significant addition to the explanatory 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 $\tau_4 = -0.0171$.**
- **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%.**

Quality Adjustment for the Nearness to Subway Lines and Subway Travel Time: Results (cont)

- **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. These are sensible results!**
- **The estimated time from the nearest subway station to the Tokyo station parameters were $\mu_2 = -0.0008$, $\mu_3 = -0.0128$ and $\mu_4 = -0.0188$.**
- **Thus 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%,; 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%.**

Quality Adjustment for the Nearness to Subway Lines and Subway Travel Time: Results (cont)

- **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. There is little change from our previous Charts.**



Allowing for Land and Structure Price Differences Across Wards

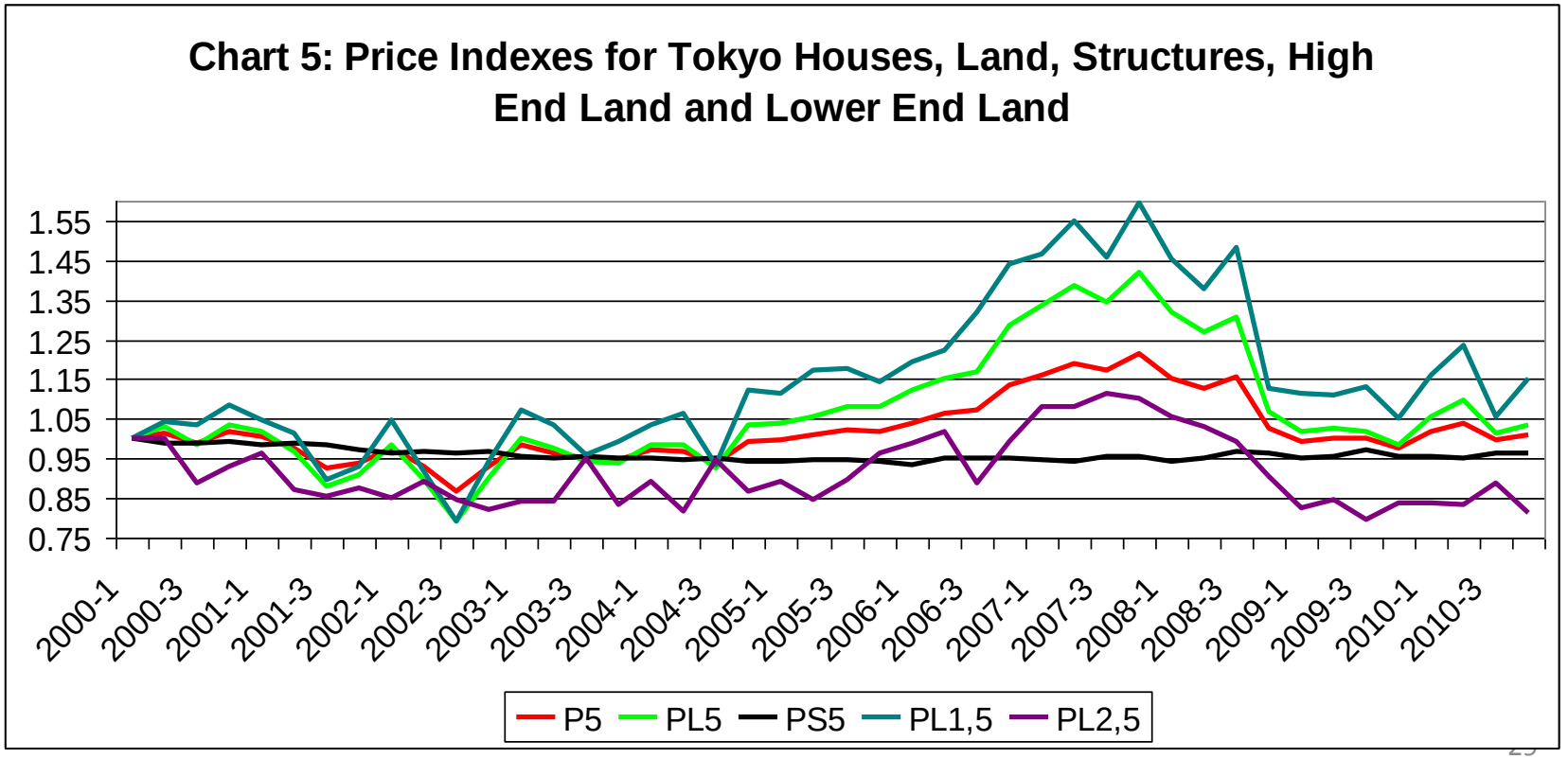
- **Usually, land price movements in high end properties are more volatile than in lower end properties.**
- **It would be preferable to have separate land price parameters (the α_t) for each Ward. However, we do not have enough degrees of freedom to accurately measure land price movements ward by ward. 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.**

Allowing for Land and Structure Price Differences Across Wards (cont)

- **We also allow for separate lot size quality adjustments in the high and lower end wards.**
- **Finally, we 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 hedonic regression model (Model 5) is defined by equations (38) in the paper along with the normalizations in equations (39).**
- **There are 128 unknown parameters to be estimated in Model 5.**

Allowing for Land and Structure Price Differences Across Wards: Results

- The R^2 for Model 5 was 0.8476 and the log likelihood was –8709.9, an increase of 106.0 over the Model 4 log likelihood.
- The resulting overall house price index P5, the overall land price PL5, the land price indexes for high and low end Wards, PL1 and PL2 respectively are plotted on Chart 5.

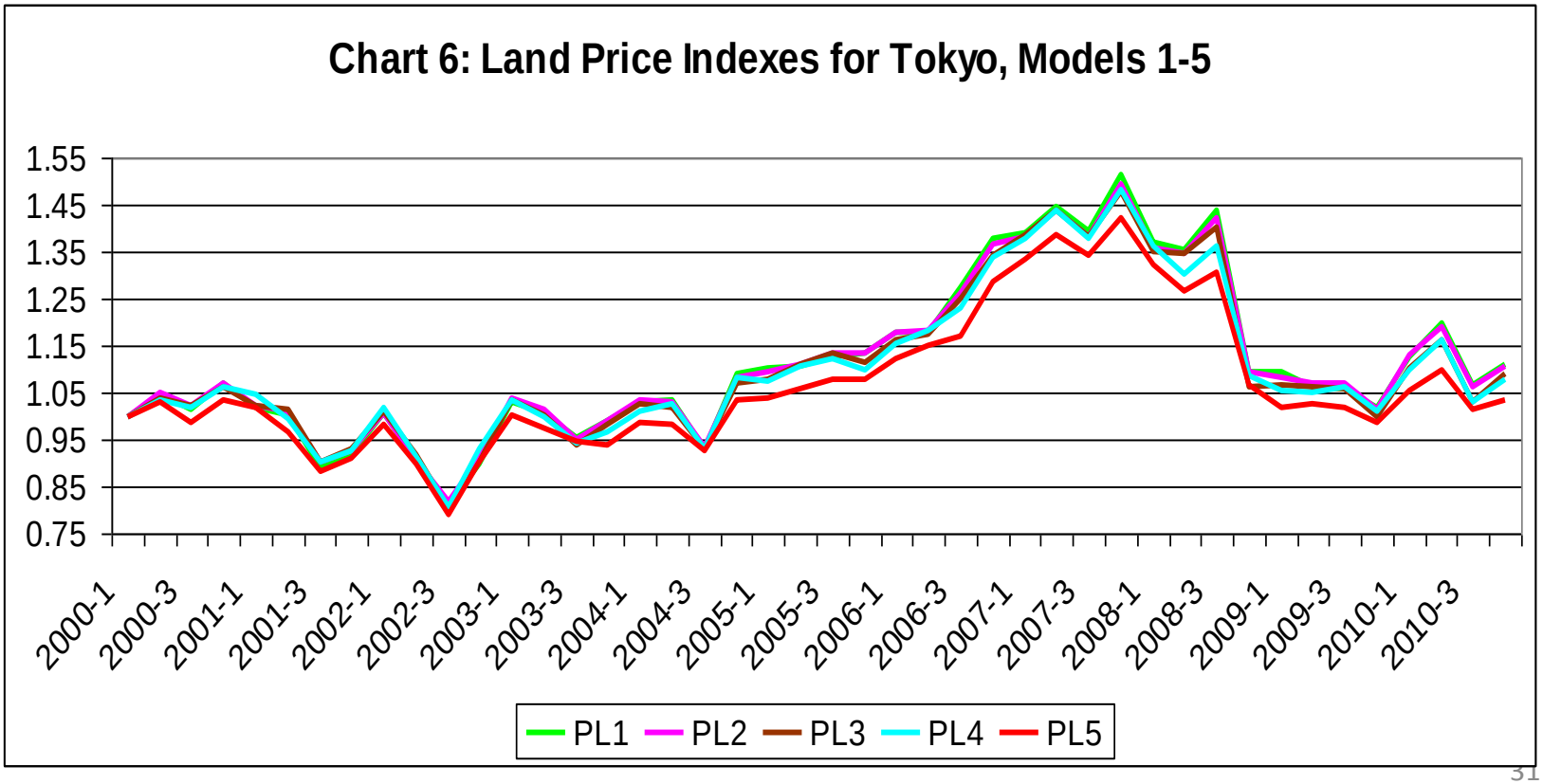


Allowing for Land and Structure Price Differences Across Wards: Results (cont)

- 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 $P_{L2,5t}$.
- 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.
- We compare the Model 1 to Model 5 overall land price indexes in Chart 6 on the following slide.

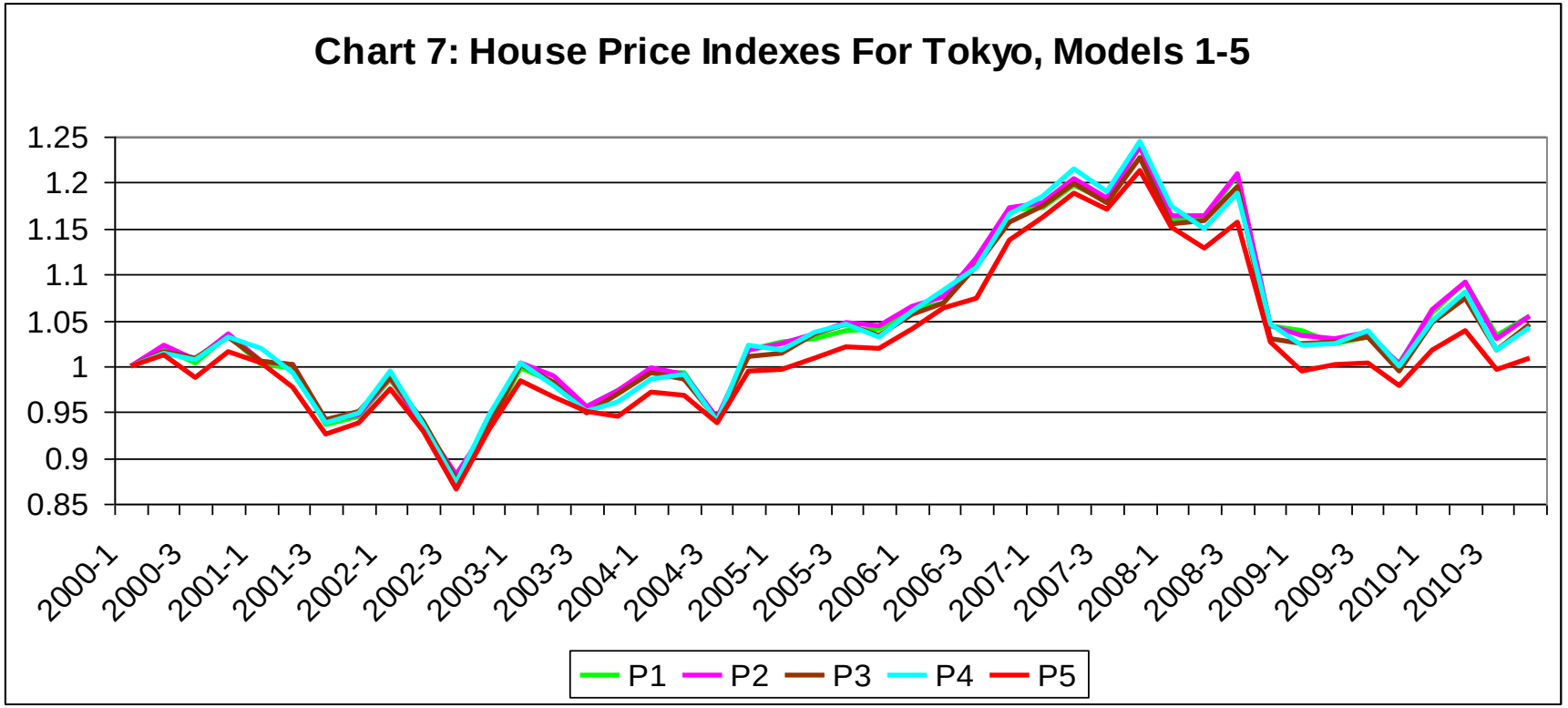
Comparison of Overall Land Price Indexes, Models 1-5

- **It can be seen that the overall land price series for Models 1-4, P_{L1t}-P_{L4t}, are generally quite close .The overall land price series for Model 5 drops a more substantial amount: about 3% on average from the other series.**



Comparison of Overall House Price Indexes, Models 1-5

- **It is also useful to compare the overall house price indexes for Models 1-5 and this is done in Chart 7 below.**
- **The Model 5 overall house price index P_{5t} is about 2% lower on average compared to the levels in the other Models.**

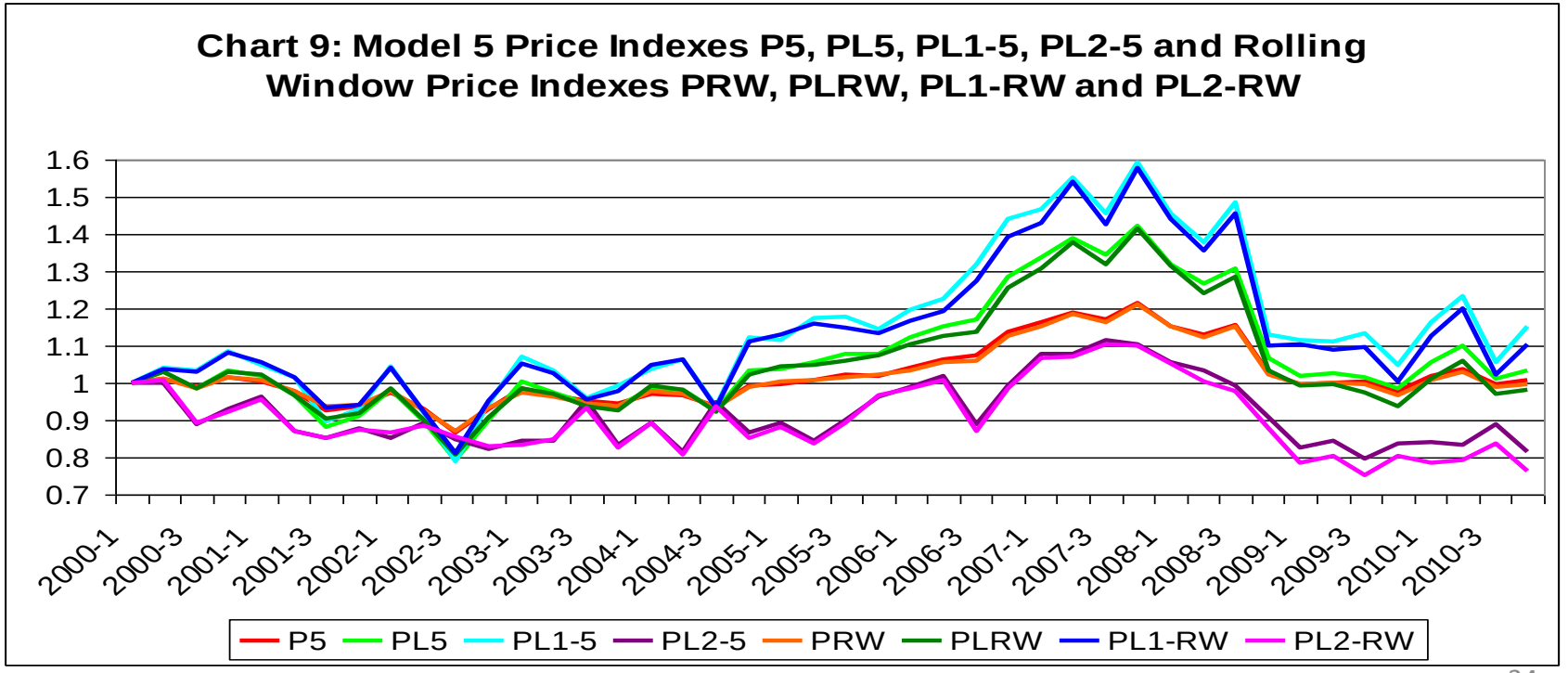


Rolling Window Hedonic Housing Regressions

- **Our Rolling Window regressions worked as follows: 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 using Model 5 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 Model 5 on the new data set. Using these new coefficient estimates, we computed the structure price index and land price indexes for high and low end wards 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.**
- **And so on until we reached Quarter 44. Thus we ran a total of 21 separate Rolling Window hedonic regressions**

Rolling Window Hedonic Housing Regressions: Results

- **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 along with their Model 5 counterpart indexes, P_{5t}, P_{L5t}, $P_{L1,5t}$ and $P_{L2,5t}$.**



Rolling Window Hedonic Housing Regressions: Results (cont)

- **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.**
- **Which set of results do we prefer? 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.**

Conclusion

- **The Generalized Builder's Model that we developed in this paper (Model 5) worked well for a large urban center (Tokyo) where the data was relatively sparse; i.e., we were able to generate sensible land and structure price indexes for Tokyo using quarterly data on house sales.**
- **Our spline approach to modeling the effects on price of various characteristics generated economically sensible estimates for the characteristics prices (with the possible exception of the number of bedrooms variable).**
- **Our Rolling Year approach also worked well and thus it would be a suitable approach for statistical agencies that are obligated to produce indexes that are not revised.**
- **An open question is: what is the "optimal" length of the window in the Rolling Window approach? This topic needs further research.**