Residential Property Price Indexes: Spatial Coordinates versus Neighbourhood Dummy Variables

W. Erwin Diewert and Chihiro Shimizu,[1] September 1, 2019
Discussion Paper 19-09,
Vancouver School of Economics,
The University of British Columbia,
Vancouver, Canada, V6T 1Z1.

Abstract

The paper addresses the following question: can satisfactory residential property price indexes be constructed using hedonic regression techniques where location effects are modeled using local neighbourhood dummy variables or is it necessary to use spatial coordinates to model location effects. Hill and Scholz (2018) addressed this question and found, using their hedonic regression model, that it was not necessary to use spatial coordinates to obtain satisfactory property price indexes for Sydney. However, their hedonic regression model did not estimate separate land and structure price indexes for residential properties. In order to construct national balance sheet estimates, it is necessary to have separate land and structure price indexes. The present paper addresses the Hill and Scholz question in the context of providing satisfactory residential *land* price indexes. The spatial coordinate model used in the present paper is a modification of Colwell's (1998) spatial interpolation method. The modification can be viewed as a general nonparametric method for estimating a function of two variables.

Key Words

Residential property price indexes, System of National Accounts, Balance Sheets, methods of depreciation, land and structure price indexes, hedonic regressions, spatial coordinates, bilinear interpolation methods, nonparametric methods for fitting two dimensional surfaces.

Journal of Economic Literature Classification Numbers

C2, C14, C21, C23, C25, C43, 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: Center for Spatial Information Science, The University of Tokyo and Nihon University (e-mail: cshimizu@csis.u-tokyo.ac.jp) . The authors thank Jan de Haan for helpful comments and gratefully acknowledge financial support from the SSHRC of Canada and Nomura Foundation.

1. Introduction

It is a difficult task to construct constant quality price indexes for residential (and commercial) properties. Properties with structures on them consist of two main components: the land component and the structure component. The problem is that each property has a unique location (which affects the price of the land component) and given the fact that the same property is not sold in every period, it is difficult to apply the usual matched model methodology when constructing constant quality price indexes. Bailey, Muth and Nourse (1963) developed the repeat sales methodology in an attempt to apply the matched model methodology to the problem of constructing property price indexes but this methodology does not allow for the use of single sales of the same property over the sample period and thus in particular, the sales of properties with new structures do not affect the resulting indexes, which could lead to biased indexes. Moreover, properties with structures on them do not retain the same quality over time due to structure depreciation and renovations or additions to the structure. Thus the matched model methodology for the construction of constant quality price indexes does not work in the property price index context.

A possible solution to the above measurement problem is to use a hedonic regression model approach to the construction of property price indexes.[2] This approach regresses the sale price of a property (or the logarithm of the sale price) on various characteristics of the properties in the sample. An important price determining characteristic of a property is its location. The location of a property can be described by its neighbourhood (a local government area or a postal code) or by its latitude and longitude, the spatial coordinates of the property. Most hedonic property regressions use the former approach to describing the location of a property but in recent years, the availability of spatial coordinate information has grown. Colwell (1998) was an early pioneer in the use of spatial coordinate information in a property price regression and more recently, Hill and Scholz (2018) used spatial coordinates to model Sydney house prices.

The main question that this paper addresses is the following one: can satisfactory residential property price indexes be constructed using hedonic regression techniques where location effects are modeled using local neighbourhood dummy variables or is it necessary to use spatial coordinates to model location effects. Hill and Scholz (2018) addressed this question and found that it was not necessary to use spatial coordinates to obtain satisfactory property price indexes for Sydney. However, their hedonic regression model did not estimate separate land and structure price indexes for residential properties. In order to construct national balance sheet estimates, it is necessary to have separate land and structure price indexes. The present paper addresses the Hill and Scholz question in the context of providing satisfactory residential *land* price indexes. The spatial coordinate model used in the present paper is a modification of Colwell's (1998) spatial interpolation method. The modification can be viewed as a general nonparametric method for estimating a function of two variables.

[2] For expositions of the hedonic regression approach to the construction of constant quality price indexes, see de Haan and Diewert (2013) and Hill (2013).

A basic building block in Colwell's method is a method of *bilinear interpolation* over a square that was developed in the mathematics literature. We explain this method in section 2 below.

In section 3, we explain how this bilinear method of interpolation over a square can be extended to a method of interpolation over a grid of squares. We then follow the example of Poirier (1976) and Colwell (1998) and convert the interpolation method into an econometric estimation model. The resulting method will be used in later sections to model the land price of a property as a function of its spatial coordinates.

In section 4, we compare Colwell's spatial coordinate model with the penalized least squares approach used by Hill and Scholz (2018) in their study of Sydney property prices. We note some problems with the Hill and Scholz approach.

Section 5 describes our data on sales of residential properties in Tokyo over the 44 quarters starting in first quarter of 2000 and ending in the last quarter of 2010. We used the same data as we used in Diewert and Shimizu (2015a) on sales of residential houses in Tokyo except the present study added additional data on sales of residential properties with no structures on the land plot.

Section 6 sets out the *builder's model* approach to hedonic property price regressions. This approach uses the property's sale price as the dependent variable and splits up property value into the sum of the land and structure components. This additive decomposition approach has a long history in the property hedonic regression literature but what is relatively recent is the use of an exogenous construction cost price series to value the structure component of the decomposition. It is this use of an exogenous index that allows us to decompose property value into plausible land and structure components.[3] This section uses both the nonparametric spatial coordinate approach due to Colwell as well as the neighbourhood approach to model the influence of location on land prices. We look at the resulting land price indexes as we increase the size of the grid and we find that there is little change in these land price indexes over a reasonable range of alternative grid sizes. Section 7 adds more characteristics to the model and again looks at how the resulting land prices change as we add more characteristics.

Section 8 compares the overall property price indexes generated by the important models explained in the previous sections (instead of comparing just the land price components of residential property sales). For comparison purposes, we also compared our "best" model results with a "traditional" hedonic property price hedonic regression which regresses the logarithm of property price on a linear combination of the property characteristics and time dummy variables.[4] This traditional approach does not generate reasonable subindexes for land and structures but it can generate reasonable results for an overall property price index.

[3] The basic idea of using an exogenous cost index can be found in Diewert (2010; 33-35). See also Diewert, de Haan and Hendricks (2015).
[4] This traditional hedonic regression approach can be traced back to Court (1939).

Section 9 concludes. An Appendix contains the results of selected regression models as well as the data underlying the charts in the main text.

2. Bilinear Interpolation on the Unit Square

Our task in this section is to explain how a particular method of bilinear interpolation works for functions of two variables defined on the unit square. This method of interpolation is a basic building block that can be used to construct a method for approximating a function of two variables that is defined over the unit square. Suppose that f(x,y) is a continuous function of two variables, x and y, where $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Suppose that f takes on the values γ_{ij} at the corners of the unit square; i.e., we have:

(1) $\gamma_{00} \equiv f(0,0)$; $\gamma_{10} \equiv f(1,0)$; $\gamma_{01} \equiv f(0,1)$; $\gamma_{11} \equiv f(1,1)$.

Assuming that we know (or can estimate) the heights of the function at the corners of the unit square, we look for an approximating continuous function that satisfies counterparts to equations (1) at the corners of the unit square and is a linear function along the four line segments that make up the boundary of the unit square. Colwell (1998; 89) showed that the following quadratic function of x and y, g(x,y), satisfies these requirements:[5]

(2) $g(x,y) \equiv \gamma_{00}(1-x)(1-y) + \gamma_{10}x(1-y) + \gamma_{01}(1-x)y + \gamma_{11}xy$.

Colwell (1998; 89) also showed that g(x,y) is a weighted average of γ_{00}, γ_{10}, γ_{01} and γ_{11} for (x,y) belonging to the unit square.[6] In order to gain more insight into the properties of g(x,y), rewrite g(x,y) as follows:

(3) $g(x,y) = \gamma_{00} + (\gamma_{10} - \gamma_{00})x + (\gamma_{01} - \gamma_{00})y + [(\gamma_{00} + \gamma_{11}) - (\gamma_{01} + \gamma_{10})]xy$.

Thus if $\gamma_{00} + \gamma_{11} = \gamma_{01} + \gamma_{10}$, then g(x,y) is a linear function over the unit square. However, if $\gamma_{00} + \gamma_{11} \neq \gamma_{01} + \gamma_{10}$, then g(x,y) is a saddle function; i.e., the determinant of the matrix of second order partial derivatives of g(x,y), $\nabla^2 g(x,y)$, is equal to $-[(\gamma_{00} + \gamma_{11}) - (\gamma_{01} + \gamma_{10})]^2 < 0$ and hence $\nabla^2 g(x,y)$ has one positive and one negative eigenvalue.

In the following section, we will follow the example of Colwell (1998; 91) and show how the function g(x,y) defined over the unit square can be extended in order to define a continuous function over a grid of squares.

3. Bilinear Spline Interpolation over a Grid

[5] The function g(x,y) defined by (2) is a special case of the bilinear function defined in matrix algebra textbooks such as Mirsky (1955; 353). Poirier (1976; 61) also defined the counterpart to (2) that is defined over a rectangle. The extension of our algebra to a grid of rectangles is straightforward.

[6] It is straightforward to show that the sum of the nonnegative weights $(1-x)(1-y)$, $x(1-y)$, $(1-x)y$ and xy is equal to 1. Thus g(x,y) will satisfy the inequalities $\min\{\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11}\} \leq g(x,y) \leq \max\{\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11}\}$.

In order to explain how Colwell's method works over a grid of squares, we will explain his method for the case of a 3 by 3 grid of squares. The method will be applied to the variables X and Y that are defined over a rectangular region in X,Y space. We assume that X and Y satisfy the following restrictions:

(4) $X_{min} \leq X \leq X_{max}$; $Y_{min} \leq Y \leq Y_{max}$

where $X_{min} < X_{max}$ and $Y_{min} < Y_{max}$. We translate and scale X and Y so that the range of the transformed X and Y, x and y, lie in the interval joining 0 and 3; i.e., define x and y as follows:

(5) $x \equiv 3(X - X_{min})/(X_{max} - X_{min})$; $y \equiv 3(Y - Y_{min})/(Y_{max} - Y_{min})$.

Define the following 3 *dummy variable* (or indicator) *functions* of x:

(6) $D_1(x) \equiv 1$ if $0 \leq x < 1$; $D_1(x) \equiv 0$ if $x \geq 1$;
$D_2(x) \equiv 1$ if $1 \leq x < 2$; $D_2(x) \equiv 0$ if $x < 1$ or $x \geq 2$;
$D_3(x) \equiv 1$ if $2 \leq x \leq 3$; $D_3(x) \equiv 0$ if $x < 2$.

Note that if $0 \leq x \leq 3$, then $D_1(x) + D_2(x) + D_3(x) = 1$ so that the 3 dummy variable functions sum to 1 if x lies in the interval between 0 and 3.

The above definitions can be used to define the 3 *dummy variable functions* of y, $D_1(y)$, $D_2(y)$ and $D_3(y)$, where y replaces x in definitions (6). Finally, a set of $3 \times 3 = 9$ *bilateral dummy variable functions*, $D_{ij}(x,y)$, is defined as follows:

(7) $D_{ij}(x,y) \equiv D_i(x)D_j(y)$; i = 1,2,3; j = 1,2,3.

The domain of definition for the $D_{ij}(x,y)$ is the *square* S_3 in two dimensional space with each side of length 3; i.e., $S_3 \equiv \{ (x,y) : 0 \leq x \leq 3; 0 \leq y \leq 3\}$. Note that for any (x,y) belonging to S_3, we have $\Sigma_{i=1}^{3} \Sigma_{j=1}^{3} D_{ij}(x,y) = 1$. Thus the bilateral dummy variable functions $D_{ij}(x,y)$ will allocate any $(x,y) \in S_3$ to one of the nine unit square cells that make up S_3. Denote the *cell* of area 1 that corresponds to x and y such that $D_{ij}(x,y) = 1$ as C_{ij} for i,j = 1,2,3. Thus the 3 cells in the grid of 9 cells that correspond to y values that satisfy $0 \leq y < 1$ are C_{11}, C_{21} and C_{31}. The 3 cells that correspond to y values such that $1 \leq y < 2$ are C_{12}, C_{22} and C_{32} and the 3 cells that correspond to y values such that $2 \leq y \leq 3$ are C_{13}, C_{23} and C_{33}.

Let f(x,y) be the function defined over S_3 that we wish to approximate. Define the *heights* γ_{ij} of the function f(x,y) at the 16 vertices of the grid of unit area cells as follows:

(8) $\gamma_{ij} \equiv f(i,j)$; i = 0,1,2,3; j = 0,1,2,3.

Define the Colwell (1998; 91-92) *bilinear spline interpolating approximation* $g_3(x,y)$ to f(x,y) for any $(x,y) \in S_3$ as follows:

$$
\begin{aligned}
(9)\ g_3(x,y) \equiv\ & D_{11}(x,y)[\phi_{00}(1-x)(1-y)+\phi_{10}(x-0)(1-y)+\phi_{01}(1-x)(y-0)+\phi_{11}xy] \\
& + D_{21}(x,y)[\phi_{10}(2-x)(1-y)+\phi_{20}(x-1)(1-y)+\phi_{11}(2-x)(y-0)+\phi_{21}xy] \\
& + D_{31}(x,y)[\phi_{20}(3-x)(1-y)+\phi_{30}(x-2)(1-y)+\phi_{21}(3-x)(y-0)+\phi_{31}xy] \\
& + D_{12}(x,y)[\phi_{01}(1-x)(2-y)+\phi_{11}(x-0)(2-y)+\phi_{02}(1-x)(y-1)+\phi_{12}xy] \\
& + D_{22}(x,y)[\phi_{11}(2-x)(2-y)+\phi_{21}(x-1)(2-y)+\phi_{12}(2-x)(y-1)+\phi_{22}xy] \\
& + D_{32}(x,y)[\phi_{21}(3-x)(2-y)+\phi_{31}(x-2)(2-y)+\phi_{22}(3-x)(y-1)+\phi_{32}xy] \\
& + D_{13}(x,y)[\phi_{02}(1-x)(3-y)+\phi_{12}(x-0)(3-y)+\phi_{03}(1-x)(y-2)+\phi_{13}xy] \\
& + D_{23}(x,y)[\phi_{12}(2-x)(3-y)+\phi_{22}(x-1)(3-y)+\phi_{13}(2-x)(y-2)+\phi_{23}xy] \\
& + D_{33}(x,y)[\phi_{22}(3-x)(3-y)+\phi_{32}(x-2)(3-y)+\phi_{23}(3-x)(y-2)+\phi_{33}xy].
\end{aligned}
$$

It can be verified that $g_3(x,y)$ is a continuous function of x and y over S_3 and $g_3(x,y)$ is equal to the underlying function f(x,y) when (x,y) is a vertex point of the grid; i.e., we have the following equalities for the 16 vertex points in S_3:

$$(10)\ g_3(i,j) = \gamma_{ij} \equiv f(i,j);\ i = 0,1,2,3;\ j = 0,1,2,3.$$

For each square of unit area in the grid, it can be seen that $g_3(x,y)$ behaves like the bilinear interpolating function g(x,y) that was defined by (2) in the previous section. Thus if (x,y) belongs to the cell C_{ij} where i and j are equal to 1, 2 or 3, then $g_3(x,y)$ is bounded from below by the minimum of the 4 vertex point values $\gamma_{i-1,j-1}$, $\gamma_{i,j-1}$, $\gamma_{i-1,j}$, $\gamma_{i,j}$ and bounded from above by the maximum of the 4 vertex point values $\gamma_{i-1,j-1}$, $\gamma_{i,j-1}$, $\gamma_{i-1,j}$, $\gamma_{i,j}$.

Following Colwell (1998; 89), if we set y = j where j = 0, 1, 2 or 3, then the resulting function of x, $g_3(x, j)$, is a *linear spline function in x* between 0 and 3; i.e., $g_3(x, j)$ is a continuous, piecewise linear function of x that has 3 (joined) linear segments that can change their slopes at the break points x = 1 and x = 2. Similarly, if we set x = i where i = 0, 1, 2 or 3, then the resulting function of y, $g_3(i, y)$, is also a *linear spline function in y* between 0 and 3. Thus we can view $g_3(x,y)$ as an interpolating function that merges these linear spline functions in the x and y directions into a consistent continuous function of two variables, where the interpolating function is equal to the function of interest at the 16 vertex points of the grid.

Following Poirier (1976; 11-12) and Colwell (1998), we can move from the interpolation model defined by (9) to an econometric estimation model. Thus suppose that we can observe x and y for N observations, say (x_n,y_n) for n = 1,...,N. Suppose also that we can observe $f(x_n,y_n)$ for n = 1,...,N. Finally, suppose that we can approximate the function f(x,y) by $g_3(x,y)$ over S_3. Let $\gamma \equiv [\gamma_{00}, \gamma_{10},...,\gamma_{33}]$ be the vector of the 16 γ_{ij} which appear in (9) and rewrite $g_3(x,y)$ as $g_3(x,y,\gamma)$. Now view γ as a vector of parameters which appear in the following linear regression model:

$$(11)\ z_n = g_3(x_n,y_n,\gamma) + \varepsilon_n\ ; \qquad n = 1,...,N.$$

If we are willing to assume that the approximation errors ε_n are independently distributed with 0 means and constant variances, the unknown parameters γ_{ij} in (11) (which are the

heights of the "true" function f(x,y) at the vertices in the grid) can be estimated by a least squares regression. It can be seen that this method for fitting a two dimensional surface over a bounded set is essentially a nonparametric method. If the number of observations N is sufficiently large and the observations are more or less uniformly distributed over the grid, then we can make the grid finer and finer and obtain ever closer approximations to the true underlying function.[7]

To see how this nonparametric approach to the estimation of a surface could be applied in the context of sales of land plots in a geographical area, suppose that in a particular time period, we have information on the selling price of N land plots. Suppose that the selling price of land plot n is P_n and the area of the property is L_n square meters. Suppose also that we have data on the latitude and longitude of property n, X_n and Y_n for n = 1,...,N. Translate and scale these spatial coordinates into the variables x_n and y_n using definitions (4) and (5) above. We suppose that N is large enough and the observations are dispersed through all 9 cells in the 3 by 3 geographical grid. An approximation to the *true land price surface* in the geographical area under consideration (which gives the price of land per meter squared as a function of the transformed spatial coordinates) can be generated by estimating the following linear regression model:

$$(12)\quad P_n/L_n = g_3(x_n,y_n,\gamma) + \varepsilon_n \;; \qquad n = 1,...,N$$

where the $g_3(x_n,y_n,\gamma)$ are defined by (9) for each (x_n,y_n) in the sample of observations. Thus estimates for the 16 unknown height parameters γ_{ij} in equations (12) can be obtained by solving a simple least squares minimization problem.

If observations are plentiful, then the grid can be made finer. Thus the 3 by 3 grid could be replaced by a k by k grid where k is an arbitrary positive integer. In this case, definitions (5) are replaced by $x \equiv k(X - X_{min})/(X_{max} - X_{min})$ and $y \equiv k(Y - Y_{min})/(Y_{max} - Y_{min})$. Definitions (6) to (9) can readily be modified to define the approximating function $g_k(x,y,\gamma)$ in place of $g_3(x,y,\gamma)$. Of course the new parameter vector γ in $g_k(x,y,\gamma)$ will have dimension $(k+1)^2$ in place of the parameter vector γ in $g_3(x,y,\gamma)$ which had dimension $4^2 = 16$. Thus Colwell (1998) realized that the well known bilinear interpolation function g(x,y) defined by (2) could be used as a basic building block in a powerful nonparametric method for approximating an arbitrary continuous function of two variables.[8]

---

[7] If the dependent variable is observed with random errors, then the method for fitting the surface can also be regarded as a smoothing method. The smoothing parameter is the number of cells in the grid, k^2 (or k can be used as the smoothing parameter); the smaller the number of cells, the smoother will be the estimated $g_k(x,y)$ function. For a discussion of smoothing methods and alternative smoothing parameters, see Buja, Hastie and Tibshirani (1989).

[8] Colwell (1998; 87) summarized his method as follows: "A simple, non-parametric approach is needed—one that fits any function with the fewest possible restrictions. The purpose of this article is to describe a method for using a single, standard OLS regression to estimate a continuous price function in space that can approximate any shape. The cost of the method developed here is found in terms of degrees of freedom. It achieves flexibility by requiring large numbers of observations." Colwell (1998; 88) after noting that his approximating function was differentiable in the interior of each square in the grid but not necessarily at boundary points of each square offered the following view on the importance of continuity versus differentiability: "This tradeoff of continuity for differentiability is worth accepting because continuity is

However, Colwell did not exhibit the explicit representation for $g_3(x,y)$ defined by (9) so it is not clear exactly how he defined his linear regression model. Colwell (1998; 92) also made the following statement about his method of parameterization: "As indicated earlier, one of the location variables must be omitted if perfect multicollinearity is to be avoided. Finally, it is not necessary to have data points within every section." Thus he seemed to suggest that one of the γ_{ij} on the right hand side of (9) needed to be omitted in order to avoid perfect multicollinearity. But such an omission would seem to destroy the flexibility of his method; i.e., setting say $\gamma_{ij} = 0$ means that we would no longer have $g_k(i,j) = f(i,j)$. Moreover, as we shall see later in our empirical application of his method, problems can arise if some cells have no observations. Thus although the spirit of his model is clear, the exact details on how to implement it are not spelled out in his paper.[9]

4. Colwell's Nonparametric Method versus Penalized Least Squares

It is useful to compare the nonparametric method for estimating a function of two variables explained in the previous section with the nonparametric method used by Hill and Scholz (2018) in their property price regressions for Sydney Australia. These authors used a penalized least squares approach for their nonparametric method.

Using the notation surrounding (11) above, a simplified version of this approach works as follows: find a function g(x,y) which is a solution to the following *penalized least squares minimization problem*:

$$(13)\ \min_g \Sigma_{n=1}^N [z_n - g(x_n,y_n)]^2 + \lambda J(g)$$

where it is assumed that g(x,y) is twice continuously differentiable and J(g) is some function of the second order partial derivatives of g evaluated at the N observed (x_n,y_n).[10] The positive parameter λ trades off how well each $g(x_n,y_n)$ approximates the observed z_n with how variable g is.

There is an extensive literature on solving this problem which is quite complicated.[11] In order to illustrate some of the problems associated with this penalized least squares approach, we will consider a simplified one dimensional version of this approach using *finite differences* of g(x) in J(g) in place of partial derivatives of g. For simplicity, we will also assume that the x_n data are unique and we order the N observations on x as $x_1 <$

---

compelling, whereas the worth of differentiability is dubious. Continuity of the price function is important because markets produce continuity. Discontinuities are destroyed by arbitrage." We agree with his assessment that differentiability of the approximating surface is not essential.

[9] Poirier (1976; 59-62) developed an approach which is equivalent to our Colwell based approach except that his parameterization of the approximating function is in terms of changes in levels rather than in the levels themselves. Thus his interaction terms are difficult to interpret. He also did not deal with the difficulties associated with empty cells.

[10] For example, J(g) could equal $\Sigma_{n=1}^N [g_{xx}(x_n,y_n) + 2g_{xy}(x_n,y_n) + g_{yy}(x_n,y_n)]$ where $g_{xx}(x_n,y_n) \equiv \partial^2 g(x_n,y_n)/\partial x \partial y$, etc.

[11] For example, see Wahba and Wendelberger (1980), Silverman (1985; 19), Wahba (2000) and Wood (2004).

$x_2 < ... < x_N$. The corresponding observed z values are z_n for n = 1,...,N. Again, for simplicity, we assume that the x_n are equally spaced.

Set $g(x_n) = s_n$ for n = 1,...,N. Our first highly simplified version problem (13) is the following *penalized least squares minimization problem*: choose $s_1, s_2, ..., s_N$ to solve the following unconstrained minimization problem:

$$(14)\ \min_{s_1,...,s_N} \{\Sigma_{n=1}^{N} [z_n - s_n]^2 + \lambda \Sigma_{n=3}^{N} [\Delta^2 s_n]^2\}$$

where $\lambda > 0$ is a positive tradeoff parameter and the first and second order finite differences of the s_n are defined as follows:

$$(15)\ \Delta s_n \equiv s_n - s_{n-1}\ ; \qquad n = 2,3,...,N;$$
$$(16)\ \Delta^2 s_n \equiv \Delta s_n - \Delta s_{n-1}\ ; \quad n = 3,4,...,N.$$

For a given λ, (14) can readily be solved using the first order conditions for the minimization problem and a bit of linear algebra. Denote the solution to (14) as the vector $s(\lambda) \equiv [s_1(\lambda),...,s_N(\lambda)]$. Denote the vector of observed z_n as $z \equiv [z_1,...,z_N]$. As λ tends to 0, $s(\lambda)$ will tend to the observed vector z. As λ tends to plus infinity, the $s_n(\lambda)$ will tend to a linear function of n; i.e., $s_n(\lambda)$ will tend to $\alpha + \beta n$ for n = 1,...,N for some α and β. This smoothing model was originally suggested by Henderson (1924; 30). Note that this smoothing method depends on the choice of λ. The method of cross validation can be used to choose λ; see Silverman (1985; 5) for references to the literature.

Our second highly simplified version problem (13) is the following *penalized least squares minimization problem*: choose $s_1, s_2, ..., s_N$ to solve the following unconstrained minimization problem:

$$(17)\ \min_{s_1,...,s_N} \{\Sigma_{n=1}^{N} [z_n - s_n]^2 + \lambda \Sigma_{n=3}^{N} [\Delta^3 s_n]^2\}$$

where $\lambda > 0$ is again a positive tradeoff parameter between fit and the variability of the s_n and the third order finite differences of the s_n, are defined as follows:

$$(18)\ \Delta^3 s_n \equiv \Delta^2 s_n - \Delta^2 s_{n-1}\ ; \quad n = 4,5,...,N.$$

Denote the solution to (17) as the vector $s(\lambda) \equiv [s_1(\lambda),...,s_N(\lambda)]$. As λ tends to 0, $s(\lambda)$ will tend to the observed vector z. As λ tends to plus infinity, the $s_n(\lambda)$ will tend to a quadratic function of n; i.e., $s_n(\lambda)$ will tend to $\alpha + \beta n + \gamma n^2$ for n = 1,...,N for some α and β. This smoothing model was originally suggested by Whittaker (1923).

In the actuarial literature, the smoothing methods that chose the $s_n = g(x_n)$ for n = 1,...,N by solving (14) or (17) for an exogenous λ is known as the Whittaker-Henderson method

of graduation[12] and in the economics literature, using (14) to smooth a time series is known as the Hodrick-Prescott (1980) filter.

This penalized least squares approach to smoothing a series implicitly defines a series to be smooth if its higher order differences are all "small". However, Bizley (1958; 126) criticized this definition of smoothness by noting that the rather smooth exponential function, $g(x) \equiv e^x$, has derivatives and differences that never become small, no matter how high a difference is taken. This fairly compelling criticism of the penalized least squares approach has been largely ignored by the current literature.

A method for smoothing a discrete series z_n can be modeled as a mapping of the vector $z \equiv [z_1,...,z_N]$ into a "smoothed" vector $s \equiv [s_1,...,s_N]$. Let $F(z) \equiv [F_1(z),...,F_N(z)]$ be the vector valued smoothing function that transforms the "rough" z into the "smooth" s so that $s \equiv F(z)$. The function F(z) is a representation of the smoothing method. Diewert and Wales (2006; 107-110) developed a *test* or *axiomatic approach* to describe desirable properties of a smoothing method. We list their tests below along with two additional tests.[13]

Test 1; Sum Preserving Test. If $s = F(z)$, then $1_N \cdot F(z) \equiv \Sigma_{n=1}^N F_n(z) = \Sigma_{n=1}^N z_n \equiv 1_N \cdot z$ where 1_N is an N dimensional vector of ones. The test says that the sum of the values of the smoothed series should equal the sum of the values of the original series.[14]

Test 2; First Moment Preserving Test: If $s = F(z)$, then $\Sigma_{n=1}^N n s_n = \Sigma_{n=1}^N n z_n$. This test was suggested by Whittaker (1923; 68).

Test 3; Identity Test: If $z = k1_N$ where k is a scalar, then $s = F(k1_N) = k1_N$. Thus if the rough z is constant, then its smooth s reproduces this constant vector.

Test 4; The Linear Trend Test: If $z_n = \alpha + \beta n$ for n = 1,...,N where α and β are constants, then F(z) = z.

Test 5; The Quadratic Trend Test: If $z_n = \alpha + \beta n + \gamma n^2$ for n = 1,...,N where α, β and γ are constants, then F(z) = z.

Test 6; The Cubic Trend Test: If $z_n = \alpha + \beta n + \gamma n^2 + \phi n^3$ for n = 1,...,N where α, β, γ and ϕ are constants, then F(z) = z.

---

[12] In the early actuarial literature, the process of smoothing a mortality table was known as *graduating the data*; i.e., the hills and valleys of the observed "rough" series, the z_n, were to be graded into a smooth road, the smoothed series, the s_n. See Sprague (1887; 112).

[13] Diewert and Wales allowed F(z) to be a set valued function rather than a single valued function. Since many smoothing methods generate a smooth that is a solution to a minimization problem; solutions to such problems may not be unique. For simplicity, here we assume unique solutions.

[14] This test implies that the arithmetic mean of the smoothed series equals the arithmetic mean of the original series. Thus this test could also be called the *mean preserving test*. Sprague (1887; 79) suggested this test.

The last 3 tests were listed in Diewert and Wales (2006; 106). The following two tests were not listed in Diewert and Wales but they are obvious tests that are similar to Tests 4-6: if the rough is a smooth elementary function of one variable, then the smooth should be identical to the rough.

Test 7; The Exponential Trend Test: If $z_n = \alpha e^n$ for n = 1,...,N where α is a constant, then F(z) = z.

Test 8; The Logarithmic Trend Test: If $z_n = \alpha \ln(n)$ for n = 1,...,N where α is a constant, then F(z) = z.

Test 9; The Diminishing Variation Test: s = F(z) implies $s \cdot s \leq z \cdot z$ or $\Sigma_{n=1}^{N} s_n^2 \leq \Sigma_{n=1}^{N} z_n^2$.

If the smoothing method satisfies Tests 1 and 9, then the variance of the smooth cannot exceed the variance of the rough. Test 9 was proposed by Schoenberg (1946; 52).

The next test is of fundamental importance in our view and it is essentially due to Sprague. It is worth quoting him on this test:

"I now proceed to a different part of my subject and prove that it is undesirable to employ such formulas as Mr. Woolhouse's, Mr. Higham's, or Mr. Ansell's, not only because, as already mentioned, they will never entirely get rid of the irregularities in our observations, but also because they all have a tendency to introduce an error even into a regular series of numbers. I start with the proposition which I think will command universal consent, that, if we attempt to graduate a perfectly regular series of numbers, the result should be to leave it unaltered; and that, if our method of procedure alters the law of the series, and substitutes for the original series one following a different law, this proves that our method of procedure is faulty." T.B. Sprague (1887; 107-108).

It can be seen that Tests 3-8 above are essentially due to Sprague. Diewert and Wales (2006; 109) argued that a consequence of the above quotation is the following test:

Test 10; The Smoothing Invariance Test: s = F(z) implies s = F(s) so that F[F(z)] = F(z). Thus if we smooth the raw data z once and obtain the smooth s = F(z) and then if we smooth the resulting s and obtain F(s), we find that the second round of smoothing just reproduces s so that F(s) = s. Put another way, the smoothing method defined by the function F should produce a smooth series and so another round of smoothing should not change the smooth series produced by the initial use of F.

It can be shown that the Henderson smoothing method defined by the solution to (14) satisfies all of the above tests except Tests 6-8 and Test 10 and the Whittaker method defined by the solution to (17) satisfies all of the above tests except Tests 7, 8 and 10.[15] The failure of a method to pass Test 10 is, in our view, is a serious problem with the method. Unfortunately, as noted by Diewert and Wales (2006; 109), most smoothing

[15] See Diewert and Wales (2006; 107-110) who drew on the earlier work of Whittaker (1923; 67), Greville (1944; 205-211), Schonberg (1946; 52) and Buja, Hastie and Tibshirani (1989; 466).

methods fail this test. For example, of the seven main types of nonparametric smoothing models listed by Buja, Hastie and Tibshirani (1989; 456-460): (i) running mean smoothers; (ii) bin smoothers; (iii) running line smoothers; (iv) polynomial regression; (v) cubic smoothing splines (the Henderson (1924) model); (vi) regression splines with fixed knots or break points; (vii) kernel smoothers; only methods (ii), (iv) and (vi) pass Test 10. The reason why these three smoothing methods pass Test 10 is that they are *linear smoothers* that are based on linear regression models.

Thus suppose that the rough z satisfies the linear regression model, $z = X\beta + \varepsilon$ where X is an N by K matrix of exogenous variables of full rank $K \leq N$ and ε is a vector of independently distributed error terms with means 0 and constant variances. Then the least squares estimator for β is $b \equiv (X^TX)^{-1}Xz$ and the predicted z vector is the smoothed vector $s \equiv Xb = X(X^TX)^{-1}Xz = Sz$ where $S \equiv X(X^TX)^{-1}X$ is the *linear smoothing matrix* for this regression based smoothing method. Thus $F(z) \equiv Sz$ for this linear smoothing method. Any linear regression based smoothing method will satisfy Test 10 since $F[F(z)] = S[Sz] = Sz = F(z)$ for this class of methods since $SS = X(X^TX)^{-1}X\ X(X^TX)^{-1}X = X(X^TX)^{-1}X = S$. Methods (ii), (iv) and (vi) are all special cases of a linear regression based smoothing method.[16]

What we are arguing in this section is that linear regression based smoothing methods have some advantages over the penalized least squares approach used by Hill and Scholz in their study. Looking at the definition of $g_3(x,y)$ in equation (9), it can be seen that our suggested bivariate nonparametric smoothing method is a linear regression based smoothing method, where the γ_{ij} are the components of the β vector in a linear regression. Thus our suggested method inherits the useful properties of a linear regression model including satisfaction of a bivariate version of Test 10; i.e., if we smooth the raw data once and then smooth once again, we do not change the original smooth, whereas the penalized least squares approach used by Hill and Scholz does not satisfy this key test. Also, if there are no random errors in the model and if the observed z_n occur precisely at the vertex points of the chosen k by k grid, then the Colwell model will reproduce the underlying functional values; i.e., we will have $g_k(x_n,y_n) = f(x_n,y_n)$ for all n where $f(x_n,y_n)$ is the "true" function value at the nth grid point. Thus our model has an underlying flexibility which is not attained by a penalized least squares approach.

We turn now to our empirical application of the Colwell nonparametric method of surface fitting.

5. The Tokyo Residential Property Sales Data

Our basic data set consists of quarterly data on V (the selling price of a residential property in Tokyo), L (the land area of the property in square meters), S (the floor space area of the structure if any on the land plot), A (the age in years of the structure if any on

[16] If the vector of ones, 1_N is spanned by the columns of X, then the linear regression based smoothing method will satisfy Tests 1, 3, 9 and 10. If in addition, the columns of X span the linear trend vector $[1,2,...,N]^T$, then the linear smoother $Sz = X(X^TX)^{-1}Xz$ will also satisfy Tests 2 and 4. Tests 5-8 will not be satisfied unless the columns of X span the particular smooth trends that are specified in these tests.

the land plot), the location of the property (specified in terms of longitude x and latitude y and in terms of the 23 Wards or local neighbourhoods of Tokyo) and some additional characteristics to be explained below. These data were obtained from a weekly magazine, *Shukan Jutaku Joho* (Residential Information Weekly) published by Recruit Co., Ltd., one of the 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.[17] We only used the price in the final week because this can be safely regarded as sufficiently close to the contract price.[18]

After range deletions, there were a total of 5580 observations with structures on the property in our sample of sales of residential property sales in the Tokyo area over the 44 quarters covering 2000-2010.[19] In addition, we had 8493 observations on residential properties with no structure on the land plot.[20] Thus there was a total of 14,073 properties in our sample. The variables used in our regression analysis to follow and their units of measurement 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;
W = Width of the lot in 1/10 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);
X = Longitude of the property;
Y = Latitude of the property;
P_S = Construction cost for a new structure in 100,000 Yen per meter squared.

[17] 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.

[18] 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 (2016).

[19] We deleted 9.2 per cent of the observations with structures 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: $1.8 \leq V \leq 20$; $0.5 \leq S \leq 2.5$; $0.5 \leq L \leq 2.5$; $1 \leq A \leq 50$; $2 \leq NB \leq 8$; $25 \leq W \leq 90$. For properties with no structure, we set the corresponding S equal to 0.

[20] The large number of plots with no structures can be explained by the preference of Japanese buyers of residential properties to construct their own house. Thus sellers of residential properties that have a relatively old structure on the property tend to demolish the structure and sell the property as a land only property.

In addition, we have the address of each property and so we can allocate each property to one of the 23 Wards of Tokyo. This information was used to construct Ward dummy variables for each property in our sample. 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	**14073**	**6.2491**	**2.9016**	**1.8**	**20**
S	**14073**	**0.43464**	**0.5828**	**0**	**2.4789**
L	**14073**	**1.0388**	**0.3986**	**0.5**	**2.4977**
A	**14073**	**5.8231**	**9.117**	**0**	**49.723**
NB	**14073**	**1.5669**	**2.0412**	**0**	**8**
W	**14073**	**46.828**	**12.541**	**25**	**90**
TW	**14073**	**9.3829**	**4.3155**	**1**	**29**
TT	**14073**	**31.244**	**7.3882**	**8**	**48**
X	**14073**	**139.67**	**0.0634**	**139.56**	**139.92**
Y	**14073**	**35.678**	**0.0559**	**35.543**	**35.816**
P_S	**14073**	**1.7733**	**0.0294**	**1.73**	**1.85**

Thus over the sample period, the sample average sale price was approximately 62.5 million Yen, the average structure space was 43.5 m^2 (but for properties with structures, the average was 110 m^2), the average lot size was 103.9 m^2, the average age of the structure was 5.8 years (for properties with a structure, the average age was 14.7 years), the average number of bedrooms in the properties that had structures was 3.95, the average lot width was 4.7 meters, the average walking time to the nearest subway station was 9.4 minutes and the average subway travelling time from the nearest station to the Tokyo Central station was 31.2 minutes.

As is usual in property regressions using L and S as independent variables, we can expect multicollinearity problems in a simple linear regression of V on S and L.[21]

In order to eliminate a possible multicollinearity problem between the lot size L and floor space area S for properties with a structure and to make our estimates of structure value consistent with structure value estimates in the Japanese national accounts, we will assume that the value of a new structure in any quarter is equal to a Residential Construction Cost Index per m^2 for Tokyo[22] (equal to P_{St} for quarter t) times the floor space area S of the structure.

6. The Basic Builder's Model using Spatial Coordinates to Model Land Prices

[21] See Diewert (2010) and Diewert, de Haan and Hendriks (2011) (2015) for evidence on this multicollinearity problem using Dutch data.

[22] 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 were constructed from the Monthly Residential Construction Cost index for Tokyo.

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[23] 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 regression model* for period t where the α_t and β_t are the parameters to be estimated in the regression:[24]

$$(19)\ V_{tn} = \alpha_t L_{tn} + \beta_t S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

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 (19) for each period t in our sample.

The hedonic regression model defined by (19) applies to new structures. But it is likely that a model that is similar to (19) 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 $A_{tn} = A(t,n)$ and assuming a geometric depreciation model, a more realistic hedonic regression model than that defined by (19) above is the following *basic builder's model*:[25]

$$(20)\ V_{tn} = \alpha_t L_{tn} + \beta_t(1-\delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t)$$

[23] 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.

[24] 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 (2008) (2010), Rambaldi, McAllister, Collins and Fletcher (2010) and Diewert, Haan and Hendriks (2011) (2015).

[25] This formulation follows that of Diewert (2008) (2010), Diewert, Haan and Hendriks (2011) (2015), de Haan and Diewert (2013) and Diewert and Shimizu (2015a). It is a special case of Clapp's (1980; 258) hedonic regression model. For applications of the builder's model to condominium sales, see Diewert and Shimizu (2017a) and Burnett-Issacs, Huang and Diewert (2016).

where the parameter δ reflects the *net depreciation rate* as the structure ages one additional period.[26] Note that (20) is now a nonlinear regression model whereas (19) was a simple linear regression model.

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 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.[27]

As was mentioned in the previous section, we have 14,073 observations on sales of houses in Tokyo over the 44 quarters in years 2000-2010. Thus equations (20) above could be combined into one big regression and a single depreciation rate δ 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 (20) due to the multicollinearity between lot size and structure size.[28] 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). Thus if the sale of property n in period t has a new structure on it, we assume that the value of this new structure is equal to this measure of residential building costs p_{St} time the floor space area of the new structure, S_{tn}. We apply this same line of reasoning to property sales that have old structures on them as well. Thus our new builder's model replaces the parameter β_t which appears in equations (20) with the exogenous official price P_{St}. Our new model becomes the following one:

$$(21)\ V_{tn} = \alpha_t L_{tn} + P_{St}(1-\delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

Thus we have 14,073 degrees of freedom to estimate 44 land price parameters α_t and one annual geometric depreciation rate parameter δ, a total of 45 parameters. We estimated the nonlinear regression model defined by (21) for our Tokyo data set using the econometric programming package Shazam; see White (2004). The R^2 for the resulting

[26] 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.

[27] Thorsnes (1997; 101) assumed that a related supply side model held instead of equation (20). 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.

[28] See Schwann (1998), Diewert (2010) and Diewert, de Haan and Hendriks (2011) (2015) on the multicollinearity problem.

preliminary nonlinear regression *Model 0* was only 0.5545,[29] which is not very satisfactory. However, there are no location variables in Model 0.

The value of a structure of the same type and age should not vary much from location to location. However, the price of land will definitely depend on the location of the property. Thus for our next model, we assume that the per meter price of land of a property is a function f(x,y) of its spatial coordinates, x and y. Thus let x_{tn} and y_{tn} equal the normalized longitude and latitude of property n sold in period t. We will initially approximate the true land price surface f(x,y) by the 4 by 4 Colwell spatial grid function $g_4(x,y)$ defined above in section 3. If X_{tn} and Y_{tn} are the raw longitude and latitude of property n sold in period t, then define the corresponding transformed spatial coordinates as $x_{tn} \equiv 4(X_{tn} - X_{min})/(X_{max} - X_{min})$ and $y_{tn} \equiv 4(Y_{tn} - Y_{min})/(Y_{max} - Y_{min})$ and define the Colwell approximation to $f(x_{tn},y_{tn})$ as $g_4(x_{tn},y_{tn})$ using the definitions in section 3. *Model 1* is the following nonlinear regression model:

$$(22)\ V_{tn} = \alpha_t g_4(x_{tn},y_{tn},\gamma)L_{tn} + P_{St}(1-\delta)^{A(t,n)}S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

Note that the γ vector of parameters in $g_4(x_{tn},y_{tn},\gamma)$ consists of the 25 spatial grid parameters γ_{ij} where i, j = 0,1,2,3,4. Thus equations (22) contain 44 unknown period t land price parameters α_t, 25 unknown γ_{ij} spatial grid parameters and 1 depreciation rate parameter δ for a total of 70 unknown parameters. However, not all of these parameters can be estimated. If we multiply all components of γ by the positive number λ and divide all α_t by λ, it can be verified that the terms $\alpha_t\, g_4(x_{tn},y_{tn},\gamma)L_{tn}$ remain unchanged. Thus a normalization on the α_t and the γ_{ij} is required. We impose the normalization $\alpha_1 = 1$ which means that the sequence, 1, $\alpha_2,...,\alpha_{44}$, can be interpreted as an index of *residential land prices* for Tokyo for the 44 quarters in our sample, where the index is set equal to 1 in the first quarter of 2000.[30]

There are 4×4 = 16 cells C_{ij} in our grid of squares where C_{11} is the cell in the southwest corner of the grid, C_{41} is the southeast corner cell, C_{14} is the northwest corner cell and C_{44} is the northeast corner cell. It turns out that cell C_{41} has no observed property sales over the entire sample period.[31] This means that γ_{44}, the value of land per meter squared at the southeast corner of the grid, cannot be identified. Thus in addition to the normalization α

[29] 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 geometric depreciation rate was $\delta = 10.49\%$, with a T statistic of 23.3. This depreciation rate is too high to be believable. As we add more explanatory variables, we will obtain more reasonable depreciation rates.

[30] Note that the α_t shift the entire land price surface $g_4(x,y,\gamma)$ in a proportional manner over time. Thus all reasonable index numbers of the land price components of individual residential properties in Tokyo will be proportional to the estimated parameter sequence 1, $\alpha_2^*,...,\alpha_{44}^*$. This is perhaps a weakness of our model but given the nonparametric nature of our modeling of land prices, some simplifying assumptions had to be made in order to estimate all of the parameters in our model. In a real time setting, a rolling window approach would be used in order to implement our model which would allow the height parameters to change over time; see Shimizu, Nishimura and Watanabe (2010) for an example of this approach.

[31] This cell is defined as properties with normalized spatial coordinates (x,y) where x and y satisfy the restrictions $3 \le x \le 4$ and $0 \le y < 1$.

= 1, we set $\gamma_{44} = 0$ in equations (22). These normalizations will ensure that the nonlinear minimization problem associated with estimating Model 1 will have a unique solution. Thus Model 2 has 68 unknown parameters.

We used Shazam's nonlinear regression option to estimate the unknown parameters in (22). The R^2 for Model 1 turned out to be 0.7973, a huge jump from the R^2 for Model 0, which was only 0.5545. This large jump indicates the importance of including locational variables in a property regression. The log likelihood for Model 1 increased by 5524.50 points over the final log likelihood of Model 0 for adding 23 new location parameters. Since Model 0 is a special case of Model 1, this is a highly significant increase in log likelihood. The estimated geometric depreciation rate from Model 1 was 6.33 % per year (T statistic = 31.7) which is more reasonable than the Model 0 estimate of 10.49 %.

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 (22) decomposes the period t value of property n into two terms: one which involves the land area L_{tn} of the property, $\alpha_t g_4(x_{tn},y_{tn},\gamma)L_{tn}$, and another term, $P_{St}(1 - \delta)^{A(t,n)}S_{tn}$, which involves the structure area S_{tn} of the property. 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 (if $S_{tn} > 0$). 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 property n sold in period t, the corresponding constant quality quantity should be $g_4(x_{tn},y_{tn},\gamma)L_{tn}$.[32] Turning to the decomposition of the structure value of property n sold in period t, $P_{St}(1 - \delta)^{A(t,n)}S_{tn}$, into price and quantity components, we take P_{St} as the price and $(1 - \delta)^{A(t,n)}S_{tn}$ as the corresponding quantity for property n sold in quarter t.

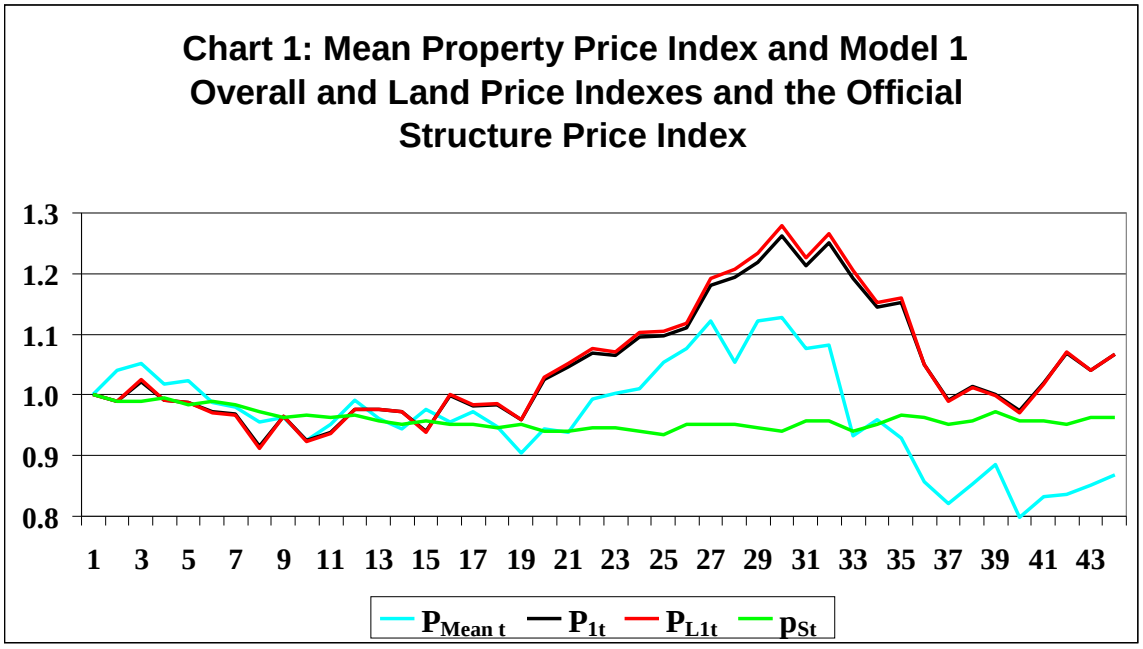
Note that the above value decompositions of individual property values into land and structure components sets the price of a square meter of land in quarter t equal to α_t^*, the estimated parameter value for α_t and sets the price of a square meter of structure equal to P_{St}, the official per meter structure cost for quarter t. These prices are assumed to be the same across all properties sold in period t and thus we can set the *aggregate* land and structure price for all residential properties sold in period t equal to P_{Lt} and P_{St} where $P_{Lt} \equiv \alpha_t^*$ for t = 1,...,44. The corresponding *aggregate* constant quality quantities of land and structures sold in period t are defined as follows:

$$(23)\ Q_{Lt} \equiv \Sigma_{n=1}^{N(t)} g_4(x_{tn},y_{tn},\gamma^*)L_{tn}\ ;\ Q_{St} \equiv \Sigma_{n=1}^{N(t)} (1 - \delta^*)^{A(t,n)}S_{tn}\ ; \qquad t = 1,...,44$$

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

where $\gamma^* \equiv [\gamma_{00}^*, ..., \gamma_{44}^*]$ and δ^* are the estimated parameter values obtained by running the nonlinear regression model defined by (22).[33]

The price and quantity series for land and structures need to be aggregated into an overall Tokyo residential property sales 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 2, P_t, as the chained Fisher price index of the land and structure series $\{P_{Lt},P_{St},Q_{Lt},Q_{St}\}$. Since these aggregate price and quantity series are generated by the Model 1 nonlinear regression model defined by equations (22), we relabel Q_{Lt}, Q_{St}, P_t, P_{Lt}, P_{St}, as Q_{L1t}, Q_{S1t}, P_{1t}, P_{L1t}, P_{S1t} for t = 1,...,T.[34]



[33] We could use hedonic imputation or index number theory to form aggregate price and quantity indexes of land and structures but because our model makes the constant quality price of land and structures the same across all property sales in a quarter, our aggregation procedure can 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).

[34] 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 $P_{Ft} \equiv [P_{LASt}P_{PAAt}]^{1/2}$ where the Laspeyres and Paasche chain link indexes are defined as $P_{LASt} \equiv [P_{L1t}Q_{L1t-1}+P_{S1t}Q_{S1t-1}]/[P_{L1t-1}Q_{L1t-1}+P_{S1t-1}Q_{S1t-1}]$ and $P_{PAAt} \equiv [P_{L1t}Q_{L1t}+P_{S1t}Q_{S1t}]/[P_{L1t-1}Q_{L1t}+P_{S1t-1}Q_{S1t}]$. 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.

The overall Model 1 house price index P_{1t} as well as the land and structure price indexes P_{L1t} and the normalized structure price index, $p_{St} \equiv P_{St}/P_{S1}$, for Tokyo over the 44 quarters in the years 2000-2010 are graphed in Chart 1.[35] We have also computed the quarterly mean selling price of properties traded in quarter t and then normalized this average property price series to start at 1 in Quarter 1 of 2000. This mean price series, $P_{Mean\ t}$, is also graphed in Chart 1. [36]

It can be seen that the official structure price series gradually trends downward over the sample period, which is not surprising since general deflation occurred in Japan during our sample period. Because there are so many land only properties in our sample and since the value of structures is relatively small for properties which have structures on them, it can be seen that our estimated land price series, P_{L1t}, is relatively close to our Model 1 overall property price index, P_{1t}. It can also be seen that the average property price series, $P_{Mean\ t}$, has the same general shape as our overall property price index P_{1t}, but the average property price series lies well below our constant quality property price series by the end of the sample period. This is to be expected since the mean property price series does not take into account depreciation of the structures for properties that have structures on them. However, the extent of the downward bias in the mean property price series by the end of the sample period is somewhat surprising.

Can we vary the number of cells in the spatial grid and explain more of the variation in residential property prices? We address this question in the next four hedonic regression models (Models 2-5) where we progressively increase the number of cells in the locational grid. Thus we will replace the land price approximating function $g_4(x_{tn},y_{tn},\gamma)$ in (22) by $g_5(x_{tn},y_{tn},\gamma)$, $g_6(x_{tn},y_{tn},\gamma)$, $g_7(x_{tn},y_{tn},\gamma)$ and $g_8(x_{tn},y_{tn},\gamma)$. The resulting Models 2-5 have 25, 36, 49 and 64 cells C_{ij} and 36, 49, 64 and 81 spatial land price height parameters γ_{ij} respectively. Setting up the corresponding nonlinear regressions using (22) as a template is straightforward except that the existence of cells with no sample observations means that not all height parameters can be estimated.

For *Model 2*, which used $g_5(x_{tn},y_{tn},\gamma)$ in (22) in place of $g_4(x_{tn},y_{tn},\gamma)$, the following cells in the 5 by 5 grid of cells had no sales over our sample period: C_{11}, C_{41}, C_{51} and C_{42}. This means that 3 height parameters could not be estimated so we imposed the following restrictions on the parameters of Model 2: $\gamma_{00} = \gamma_{40} = \gamma_{50} = 0$. We also set $\alpha_1 = 1$ so that the remaining land price parameters α_t could be identified. Thus Model 2 had $36 - 3 = 33$ γ_{ij} parameters, 43 land price parameters α_t and 1 depreciation rate parameter δ for a total of 77 parameters. The final log likelihood for Model 2 was 155.04 points higher than the final log likelihood for Model 1 for adding 9 extra land price location parameters. The resulting R^2 was 0.8035 and the estimated geometric depreciation rate was $\delta^* = 6.29\%$ with a T statistic of 31.6. We expected that all of the estimated height parameters would be positive but two of them (γ_{51}^* and γ_{05}^*) turned out to be negative. However, the

[35] Define the normalized official structure price series as $p_{St} = P_{St}/P_{S1}$ for t = 1,...,44. This is the series that is plotted in Chart 1. It will not change as we introduce additional hedonic property regression models. We note that the official index $P_{St} = 18.5p_{St}$; i.e., $P_{S1} = 18.5$.

[36] The series P_{Mean}, P_1, P_{L1} and p_S are also listed in Table A1 of the Appendix.

estimated land prices for each observation tn in our sample, $g_5(x_{tn},y_{tn},\gamma^*)$, turned out to be positive for t = 1,...,44 and n = 1,...,N(t), and so we did not worry about these 3 negative γ_{ij}^* at this stage of our investigation.[37] The sequence of estimated α_t^* is our estimated land price series for Model 2, P_{L2t}, and this series is plotted in Chart 2 below and is listed in Table A2 of the Appendix.

For *Model 3*, which used $g_6(x_{tn},y_{tn},\gamma)$ in (22) in place of $g_4(x_{tn},y_{tn},\gamma)$, the following 5 cells in the 6 by 6 grid of cells had no sales over our sample period: C_{11}, C_{51}, C_{61}, C_{52} and C_{62}. Thus we set the following 5 height parameters equal to 0 in order to identify the remaining height parameters: $\gamma_{00} = \gamma_{50} = \gamma_{60} = \gamma_{51} = \gamma_{61} = 0$. We also set $\alpha_1 = 1$ so that the remaining land price parameters α_t could be identified. Thus Model 3 had 49 – 5 = 44 γ_{ij} parameters, 43 land price parameters α_t and 1 depreciation rate parameter δ for a total of 88 parameters. The final log likelihood for Model 3 was 82.43 points *lower* than the final log likelihood for Model 2 for adding 11 extra land price location parameters. Model 3 is not a special case of Model 2 so it can happen that a moving to a larger number of squares in the grid does not improve the fit of the model. The problem is that there are likely to be discrete neighbourhood land price effects and our relatively course partition of the city into squares does not adequately capture these discrete neighbourhood effects. The resulting R^2 for Model 3 was 0.8014 (less than the Model 2 R^2 of 0.8035) and the estimated geometric depreciation rate was $\delta^* = 6.25\%$ with a T statistic of 31.8. There were 5 negative estimates for the land price height parameters: γ_{01}^*, γ_{10}^*, γ_{41}^*, γ_{06}^* and γ_{56}^*. However, the estimated land prices $g_6(x_{tn},y_{tn},\gamma^*)$ turned out to be positive for each observation in our sample. The sequence of estimated α_t^* is our estimated land price series for Model 3, P_{L3t}, and this series is plotted in Chart 2 below and is listed in Table A2 of the Appendix.

Model 4 used $g_7(x_{tn},y_{tn},\gamma)$ in (22) in place of $g_4(x_{tn},y_{tn},\gamma)$. The following 9 cells in the 7 by 7 grid of cells had no sales over our sample period: C_{11}, C_{21}, C_{51}, C_{61}, C_{71}, C_{52}, C_{62}, C_{72} and C_{17}. Thus we set the following 9 height parameters equal to 0 in order to identify the remaining height parameters: $\gamma_{00} = \gamma_{10} = \gamma_{50} = \gamma_{60} = \gamma_{70} = \gamma_{51} = \gamma_{61} = \gamma_{71} = \gamma_{07} = 0$. We also set $\alpha_1 = 1$ so that the remaining land price parameters α_t could be identified. Thus Model 4 had 64 – 9 = 55 γ_{ij} parameters, 43 land price parameters α_t and 1 depreciation rate parameter δ for a total of 99 parameters. The final log likelihood for Model 4 was 501.88 points higher than the final log likelihood for Model 3 for adding 11 extra land price location parameters. The resulting R^2 for Model 4 was 0.8156 and the estimated geometric depreciation rate was $\delta^* = 5.99\%$ with a T statistic of 31.9. There were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. As usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The sequence of estimated α_t^* is our

[37] The city of Tokyo is adjacent to the Pacific Ocean and so the boundaries of the city do not fit nicely into a rectangular grid (which we transformed into a square grid). Thus as the number of squares in the grid becomes larger, some squares at the boundaries of the grid will end up having no observations or very few observations. Thus suppose the observations in cell C_{11} are concentrated in the top north east corner of this cell. Then a better fit to the observed data in cell C_{11} may be obtained by setting γ_{00} equal to a negative number.

estimated land price series for Model 4, P_{L4t}, and this series is plotted in Chart 2 below and is listed in Table A2 of the Appendix.

Finally, *Model 5* used $g_8(x_{tn},y_{tn},\gamma)$ in (22) in place of $g_4(x_{tn},y_{tn},\gamma)$. The following 14 cells in the 8 by 8 grid of cells had no sales over our sample period: C_{11}, C_{12}, C_{21}, C_{18}, C_{61}, C_{62}, C_{63}, C_{71}, C_{72}, C_{73}, C_{81}, C_{82}, C_{83} and C_{88}. All 4 corner cells were empty along with many other boundary cells. Thus we set the following 14 height parameters equal to 0 in order to identify the remaining height parameters: $\gamma_{00} = \gamma_{10} = \gamma_{01} = \gamma_{60} = \gamma_{61} = \gamma_{62} = \gamma_{70} = \gamma_{71} = \gamma_{72} = \gamma_{80} = \gamma_{81} = \gamma_{82} = \gamma_{88} = 0$. We also set $\alpha_1 = 1$ so that the remaining land price parameters α_t could be identified. Thus Model 5 had 91 – 14 = 77 γ_{ij} parameters, 43 land price parameters α_t and 1 depreciation rate parameter δ for a total of 111 parameters. The final log likelihood for Model 5 was 249.72 points *lower* than the final log likelihood for Model 4 for adding 12 extra land price location parameters. The resulting R^2 for Model 5 was 0.8086 (compared to 0.8156 for Model 4) and the estimated geometric depreciation rate was $\delta^* = 6.18\%$ with a T statistic of 31.5. There were 5 negative estimates for the land price height parameters: γ_{02}^*, γ_{11}^*, γ_{50}^*, γ_{51}^* and γ_{52}^*. As usual, the estimated land prices, the $g_8(x_{tn},y_{tn},\gamma^*)$, turned out to be positive for each observation in our sample. The sequence of estimated α_t^* is our estimated land price series of index numbers for Model 5, P_{L5t}, and this series is plotted in Chart 2 below and is listed in Table A2 of the Appendix.

At this point, we decided stop the process of increasing the number of height parameters. It is clear that our best model up to this point was Model 4.

One of the main purposes of this paper is to see if the use of spatial coordinates in a residential hedonic property value regression can lead to more accurate estimates for a property price index and for a land price subindex for residential properties than can be obtained using just postal codes or other neighbourhood locational variables. Hill and Scholz (2018) made this comparison for residential property price indexes but not for the land price component of their overall property price index since their methodological approach did not allow for separate land and structure subindexes.

An alternative to using spatial coordinates to measure the influence of location on property prices is to use postal codes or neighbourhoods as indicators of location. There are 23 Wards in Tokyo and each property in our sample belongs to one of these Wards. 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 (20). These 23 dummy variables are defined as follows: for t = 1,...,44; n = 1,...,N(t); j = 1,...,23:[38]

(24) $D_{W,tn,j} \equiv 1$ 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.

[38]The number of observations in each Ward in our sample was as follows: 3, 5, 195, 429, 348, 28, 62, 94, 453, 1260, 1114, 3434, 382, 701, 2121, 274, 107, 76, 432, 1679, 361, 212, 303. Thus Wards 1 and 2 had very few observations.

We now modify the model defined by (20) to allow the *level* of land prices to differ across the 23 Wards of Tokyo. The new *Model 6* is defined by the following nonlinear regression model:

$$(25)\ V_{tn} = \alpha_t(\sum_{j=1}^{23} \omega_j D_{W,tn,j})L_{tn} + P_{St}(1-\delta)^{A(t,n)}S_{tn} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

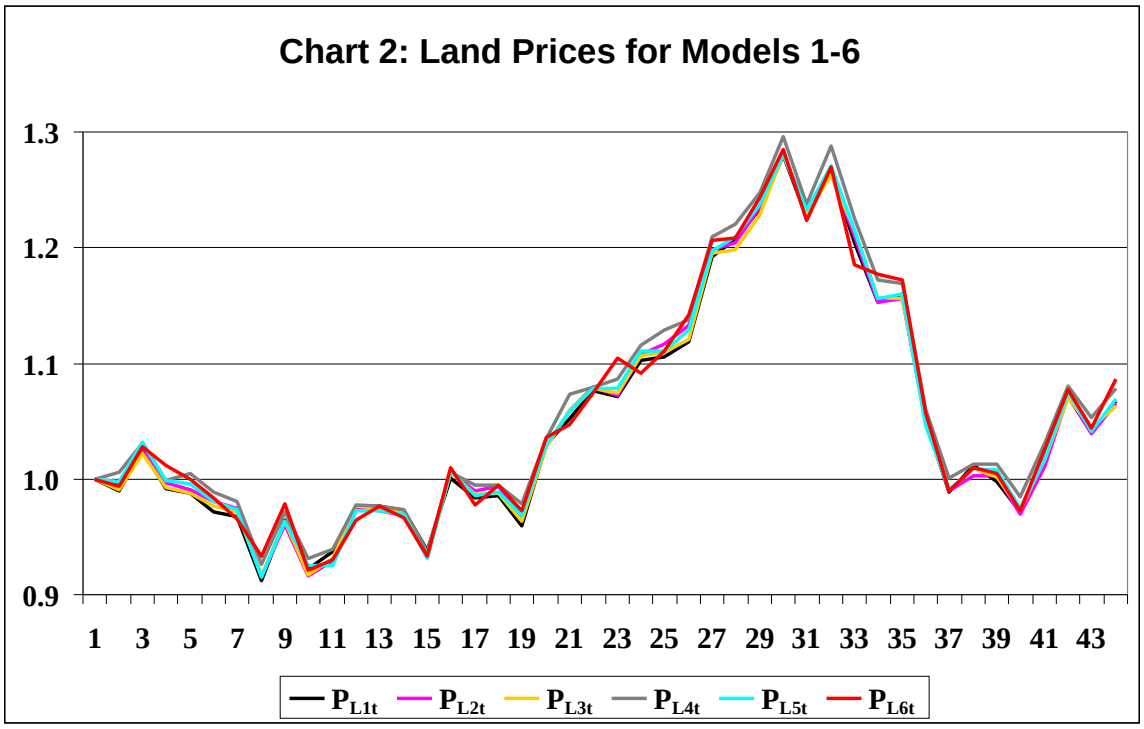
Comparing the models defined by equations (20) and (25), it can be seen that we have added an additional 23 *ward relative land value parameters*, $\omega_1,...,\omega_{23}$, to the model defined by (20). However, looking at (25), it can be seen that the 44 land price time parameters (the α_t) and the 23 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 $\alpha_1 = 1$. Thus equations (25) contain 43 unknown period t land price parameters α_t, 23 Ward relative land price parameters, the ω_j, which replace the 25 unknown γ_{ij} spatial grid parameters in (22), and 1 depreciation rate parameter δ for a total of 67 unknown parameters. Thus this *Ward dummy variable hedonic regression* (Model 6) has roughly the same number of parameters as our spatial coordinate Model 1, which had 68 unknown parameters.

The final log likelihood for Model 6 was –24318.67, a gain of 5045.90 over the final log likelihood of Model 0 defined by equations (21). The R^2 for Model 6 was 0.7853. The final log likelihood for Model 1 was –23840.07 and the R^2 was 0.7993. Thus the spatial coordinates Model 1 fit the data better than the dummy variable Model 6. Both models had roughly the same number of parameters. However, how different are the resulting land price indexes generated by these two models? As usual, the sequence of estimated α_t^* is our estimated land price series for Model 6, P_{L6t}, and this series is plotted in Chart 2 below and is listed in Table A2 of the Appendix.[39]

It can be seen that all 6 models produce much the same land price indexes.[40] Since our best fitting model was Model 4, P_{L4t} is our preferred land price series. Note that the Ward dummy variable model land price index, P_{L6t}, is fairly close to our preferred series.

[39] Define the aggregate constant quality amounts of residential land and structures sold in period t by $Q_{Lt} \equiv \Sigma_{n=1}^{N(t)} (\sum_{j=1}^{23} \omega_j^* D_{W,tn,j})L_{tn}$ and $Q_{St} \equiv \Sigma_{n=1}^{N(t)} (1-\delta^*)^{A(t,n)}S_{tn}$ for t = 1,...,44. The overall period t property price index for Model 6, P_{6t}, is defined as the chained Fisher price index using the above Q_{Lt} and Q_{St} as the period t quantity series and $P_{L6t} \equiv \alpha_t^*$ and the official structure prices P_{St} as the period t price series when constructing the Fisher index chain links.

[40] Since the structure component of overall property prices is relatively small compared to the land component and since the structure price index is the same across all 6 models, the overall property price indexes generated by Models 1-6 are all very similar.



The above 6 models make use of information on land plot size, structure floor space, the age of the structure (if the property has a structure) and its location, either in terms of spatial coordinates or terms of its neighbourhood.[41] These are the most important residential property price determining characteristics in our view. In the following section, we make use of additional information on housing characteristics and see if this extra information materially changes our estimated land price indexes.[42] We will use the spatial coordinate Model 4 as our starting point in the models which follow, since it was the best fitting model studied in this section. This model used the Colwell nonparametric model for modeling the land price surface with the 7×7 = 49 cell grid.

7. Spatial Coordinate Models that Use Additional Information

It is likely that property sales that have an older structure on the property will have a different land valuation than a nearby property of the same size that consists of cleared land, since demolition costs are not trivial. Our *Model 7* takes this possibility into account. Define the dummy variable $D_{L,tn}$ as follows for t = 1,...,44 and n = 1,...,N(t):

(26) $D_{L,tn} \equiv 1$ if observation n in period t is a land only sale;
$\equiv 0$ otherwise.

[41] We also require an exogenous building cost per square meter in order to get realistic land and structure subindexes.
[42] We are also interested in determining whether the extra information will change our estimates of structure depreciation rates.

Define $D_{S,tn} \equiv 1 - D_{L,tn}$ for t = 1,...,44; n = 1,...,N(t). Thus if property n sold in period t has a structure on it, $D_{S,tn}$ will equal 1. Model 7 estimates the following nonlinear regression:

$$(27)\ V_{tn} = \alpha_t (D_{S,tn} + \phi D_{L,tn}) g_7(x_{tn},y_{tn},\gamma) L_{tn} + P_{St}(1-\delta)^{A(t,n)} S_{tn} + \varepsilon_{tn}\ ; \quad t = 1,...,44;\ n = 1,...,N(t).$$

Thus the parameter ϕ gives the added premium to the property's land price (per meter squared) if the property has no structure on it. We expect ϕ to be a small number. Since we are using the interpolation function $g_7(x_{tn},y_{tn},\gamma)$ as a basic building block in the nonlinear regression model defined by (27), we need to impose the same restrictions on the γ_{ij} that were imposed in Model 4. Thus we set the following 9 height parameters equal to 0 in order to identify the remaining height parameters: $\gamma_{00} = \gamma_{10} = \gamma_{50} = \gamma_{60} = \gamma_{70} = \gamma_{51} = \gamma_{61} = \gamma_{71} = \gamma_{07} = 0$. We also set $\alpha_1 = 1$ so that the remaining land price parameters α_t could be identified. Thus Model 7 had 64 – 9 = 55 γ_{ij} parameters, 43 land price parameters α_t, 1 depreciation rate parameter δ and one land only premium parameter ϕ for a total of 100 parameters. As starting coefficient values for Model 7, we used the final coefficient estimates from Model 4 plus the starting value $\phi = 0$. The final log likelihood for Model 7 was 128.75 points higher than the final log likelihood for Model 4 for adding 1 land only parameter. The resulting R^2 for Model 7 was 0.8175 (the Model 4 R^2 was 0.8156) and the estimated geometric depreciation rate was $\delta^* = 2.85\%$ with a T statistic of 15.4. Recall that the estimated depreciation rate from Model 4 was 5.99%. Our new estimated depreciation rate is much more reasonable. The estimated ϕ was $\phi^* = 1.110$ (T statistic = 153.4). Thus a property without a structure sold at an 11.0% premium compared to a similar property without a structure. There were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. As usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The sequence of estimated α_t^* is our land price series for Model 7, P_{L7t}, and this series is plotted in Chart 3 below and is listed in Table A3 of the Appendix.

The most important point we learned from running this regression model is that residential property sales in Japan with and without a structure on the property are qualitatively different. Taking this difference into account led to much more reasonable estimated structure depreciation rates.

In our next model, we allow the per square meter price of land to vary as the size of the land plot increases. Recall that we have restricted the range of the land variable to $0.5 \leq L_{tn} \leq 2.5$.[43] We allow the price of land to be piecewise linear function of the plot size with 3 break points; 1, 1.5 and 2. Using these *land area break points*, we found that 7492 observations fell into the interval $0.5 \leq L_{tn} < 1$, 4711 observations fell into the interval $1 \leq L_{tn} < 1.5$, 1414 observations fell into the interval $1.5 \leq L_{tn} < 2$ and 456 observations fell into the interval $2 \leq L_{tn} \leq 2.5$. We label the four sets of observations that fall into the

[43] 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.

above four groups as groups 1-4. For each observation n in period t, we define the four *land dummy variables*, $D_{L,tn,k}$, for k = 1,2,3,4 as follows:[44]

(28) $D_{L,tn,k} \equiv 1$ 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:

$$(29)\ f_L(L_{tn},\lambda) \equiv D_{L,tn,1}[\lambda_0 L_0 + \lambda_1(L_{tn} - L_0)] + D_{L,tn,2}[\lambda_0 L_1 + \lambda_1(L_1 - L_0) + \lambda_2(L_{tn} - L_1)] + D_{L,tn,3}[\lambda_0 L_0 + \lambda_1(L_1 - L_0) + \lambda_2(L_2 - L_1) + \lambda_3(L_{tn} - L_2)] + D_{L,tn,4}[\lambda_0 L_0 + \lambda_1(L_1 - L_0) + \lambda_2(L_2 - L_1) + \lambda_3(L_3 - L_2) + \lambda_4(L_{tn} - L_3)]$$

where $\lambda \equiv [\lambda_0,\lambda_1,\lambda_2,\lambda_3,\lambda_4]$ and the λ_k are 5 unknown parameters and $L_0 \equiv 0.5$, $L_1 \equiv 1$, $L_2 \equiv 1.5$ and $L_3 \equiv 2$. The function $f_L(L_{tn},\lambda)$ defines a *relative valuation function for the land area of a house* as a function of the plot area. Thus λ_0 can be interpreted as the marginal price of land for plots between 0 and 0.5, λ_1 can be interpreted as the marginal price of land for plots between 0.5 and 1, λ_2 can be interpreted as the marginal price of land for plots between 1 and 1.5 and so on.

Model 8 is the following nonlinear regression:

$$(30)\ V_{tn} = \alpha_t (D_{S,tn} + \phi D_{L,tn}) g_7(x_{tn},y_{tn},\gamma) f_L(L_{tn},\lambda) + P_{St}(1-\delta)^{A(t,n)} S_{tn} + \varepsilon_{tn};\quad t = 1,...,44;\ n = 1,...,N(t).$$

where the function f_L is defined above by (29) and ε_{tn} is an error term. There are 44 unknown land price parameters α_t, 1 land only premium parameter ϕ, 64 land price height parameters γ_{ij}, 5 marginal price of land parameters λ_k and 1 depreciation rate δ to estimate. However, as was the case with Model 4 and the previous Model 7, some cells in our grid of cells are empty and so we cannot estimate 9 of the γ_{ij} and so there are only 55 γ_{ij} parameters to estimate. Also, we cannot identify all of the α_t and λ_k so we impose the normalizations $\alpha_1 = 1$ and $\lambda_1 = 1$. Thus we are left with 104 unknown parameters to estimate.

As starting coefficient values for Model 8, we used the final coefficient estimates from Model 7 plus the starting values $\lambda_0 = \lambda_2 = \lambda_3 = \lambda_4 = 1$. The final log likelihood for Model 8 was 328.27 points higher than the final log likelihood for Model 7 for adding 4 lot size parameters. The resulting R^2 for Model 8 was 0.8222 (the Model 7 R^2 was 0.8175) and the estimated geometric depreciation rate was $\delta^* = 3.44\%$ with a T statistic of 16.1. The estimated ϕ was $\phi^* = 1.091$ (T statistic = 155.9). Thus a property without a structure sold at an 9.1% premium compared to a similar property without a structure. There were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. As usual, the

[44] Note that for each observation, the land dummy variables sum to one; i.e., for each tn, $D_{L,tn,1} + D_{L,tn,2} + D_{L,tn,3} + D_{L,tn,4} = 1$.

estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The sequence of relative marginal valuations of land (the λ_k^*) were as follows (with the T statistics in brackets): $\lambda_0^* = 1.50$ (41.2), $\lambda_1^* = 1$ (imposed restriction), $\lambda_2^* = 1.16$ (35.1), $\lambda_3^* = 1.23$ (35.9) and $\lambda_4^* = 0.89$ (13.9). Thus as lot size increases, the per meter price of land eventually decreases. The sequence of estimated α_t^* is our index of land prices for Model 8, P_{L8t}, and this series is plotted in Chart 3 below and is listed in Table A3 of the Appendix.

In our next model, we allow the per square meter price of a square meter of structure to vary as the floor space of the structure increases. The rational for this model is that bigger houses are likely to be of higher quality. Recall that we have restricted the range of the structure floor space variable to $0.5 \leq S_{tn} \leq 2.5$.[45] We allow the price of a square meter of floor space area to be piecewise linear function of the overall floor space size with 2 break points; 1 and 1.5. Using these *structure area break points*, we found that 2768 observations fell into the interval $0.5 \leq S_{tn} < 1$, 2020 observations fell into the interval $1 \leq S_{tn} < 1.5$ and 792 observations fell into the interval $1.5 \leq S_{tn} \leq 2.5$. We label the 3 sets of observations that fall into the above 3 groups as groups 1-3. For each observation n in period t, we define the 3 *structure dummy variables*, $D_{S,tn,m}$, for m = 1,2,3 as follows:[46]

(31) $D_{S,tn,m} \equiv 1$ if observation tn has structure area that belongs to group m;
$\equiv 0$ if observation tn has structure area that does not belong to group m.

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

$$(32)\ f_S(S_{tn},\mu) \equiv D_{S,tn,1}[\mu_0 S_0 + \mu_1(S_{tn} - S_0)] + D_{S,tn,2}[\mu_0 S_1 + \mu_1(S_1 - S_0) + \mu_2(S_{tn} - S_1)] + D_{S,tn,3}[\mu_0 S_0 + \mu_1(S_1 - S_0) + \mu_2(S_2 - S_1) + \mu_3(S_{tn} - S_2)]$$

where $\mu \equiv [\mu_0,\mu_1,\mu_2,\mu_3]$ and the μ_m are 4 unknown parameters and $S_0 \equiv 0.5$, $S_1 \equiv 1$ and $S_2 \equiv 1.5$. The function $f_S(S,\mu)$ defines a *relative valuation function for the floor space area of a house* as a function of the total floor space area, S. Thus μ_0 can be interpreted as the marginal price of a structure for structures with total floor between 0 and 0.5, μ_1 can be interpreted as the marginal price of an additional square meter of structure for total structure areas between 0.5 and 1, μ_2 can be interpreted as the marginal price of an additional square meter of structure for total structure areas between 1 and 1.5 and μ_3 can be interpreted as the marginal price of an additional square meter of structure for total structure areas between 1.5 and 2.5.

Model 9 is the following nonlinear regression:

---

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

[46] Note that for each observation tn where $S_{tn} > 0$, the structure dummy variables sum to one; i.e., for each such tn, $D_{S,tn,1} + D_{S,tn,2} + D_{S,tn,3} = 1$. There were 5580 observations which had a positive S_{tn}. Note also that $f_S(S_{tn},\mu) = 0$ if $S_{tn} = 0$.

$$(33)\ V_{tn} = \alpha_t (D_{S,tn} + \phi D_{L,tn}) g_7(x_{tn},y_{tn},\gamma) f_L(L_{tn},\lambda) + P_{St}(1-\delta)^{A(t,n)} f_S(S_{tn},\mu) + \varepsilon_{tn}\ ; \quad t = 1,...,44;\ n = 1,...,N(t).$$

where the function f_L is defined above by (29), the function f_S is defined by (32) and ε_{tn} is an error term. There are 44 unknown land price parameters α_t, 1 land only premium parameter ϕ, 64 land price height parameters γ_{ij}, 5 marginal price of land parameters λ_k, 4 marginal price of structure parameters μ_m and 1 depreciation rate δ to estimate. However, as was the case with the previous Model 8, some cells in our grid of cells are empty and so we cannot estimate 9 of the γ_{ij} and so there are only 55 γ_{ij} parameters to estimate. Also, we cannot identify all of the α_t and λ_k so we impose the normalizations $\alpha_1 = 1$ and $\lambda_1 = 1$. We also impose the normalization $\mu_1 = 1$ in order to use the official structure building cost index to value new buildings with total floor space area between 0.5 and 1. This will ensure that our estimated structure values for new buildings are close to estimated structure values based solely on the official cost index. Thus we are left with 107 unknown parameters to estimate.

As starting coefficient values for Model 9, we used the final coefficient estimates from Model 8 plus the starting values $\mu_0 = \mu_2 = \mu_3 = 1$. The final log likelihood for Model 9 was 136.32 points higher than the final log likelihood for Model 8 for adding 3 structure size parameters. The resulting R^2 for Model 9 was 0.8256 (the Model 8 R^2 was 0.8222) and the estimated geometric depreciation rate was $\delta^* = 4.35\%$ with a T statistic of 18.6. The estimated ϕ was $\phi^* = 1.159$ (96.8). Thus a property without a structure sold at an 15.9% premium compared to a similar property without a structure. As usual, there were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. Also as usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The sequence of relative marginal valuations of land (the λ_k^*) were as follows: $\lambda_0^* = 1.509$ (41.7), $\lambda_1^* = 1$ (imposed restriction), $\lambda_2^* = 1.160$ (35.7), $\lambda_3^* = 1.238$ (36.8) and $\lambda_4^* = 0.898$ (14.3). The sequence of relative marginal valuations of floor space area (the μ_m^*) were as follows: $\mu_0^* = 1.461$ (21.8), $\mu_1^* = 1$ (imposed restriction), $\mu_2^* = 2.331$ (18.6) and $\mu_3^* = 1.572$ (10.0). The sequence of estimated α_t^* is our index of land prices for Model 9, P_{L9t}, and this series is plotted in Chart 3 below and is listed in Table A3 of the Appendix.

Note that the 3 models that we introduced in this section did not require any additional property characteristics: we simply made better use of the information on S and L. However, for our next model, we make use of the two subway variables: TW, the walking time in minutes to the nearest subway station, and TT, the subway running time in minutes to the Tokyo central station. The sample minimum time for TW was 1 minute and the minimum time for TT was 8 minutes. Our next model allows the price of land to decrease as these two subway time variables increase. These variables have proven to be highly significant in other studies of Tokyo property prices.[47] Thus *Model 10* is the following nonlinear regression:

[47] See for example Shimizu, Nishimura and Watanabe (2010) and Diewert and Shimizu (2015a) (2017a).

$$(34)\ V_{tn} = \alpha_t [D_{S,tn} + \phi D_{L,tn}] g_7(x_{tn},y_{tn},\gamma) f_L(L_{tn},\lambda) [1+\tau(TW_{tn}-1)][1+\rho(TT_{tn}-8)] + P_{St}(1-\delta)^{A(t,n)} f_S(S_{tn},\mu) + \varepsilon_{tn}; \quad t = 1,...,44; n = 1,...,N(t).$$

where the function f_L is defined above by (29), the function f_S is defined by (32), τ is the percentage change in the price of land due to a one minute increase in walking time, ρ is the percentage change in the price of land due to a one minute increase in subway running time to Tokyo central station and ε_{tn} is an error term. There are 44 unknown land price parameters α_t, 1 land only premium parameter ϕ, 64 land price height parameters γ_{ij}, 5 marginal price of land parameters λ_k, 4 marginal price of structure parameters μ_m, 2 subway time parameters and 1 depreciation rate δ to estimate. However, as was the case with the previous models in this section, some cells in our grid of 49 cells are empty and so we cannot estimate 9 of the γ_{ij} and so there are only 55 γ_{ij} parameters to estimate. Also, we cannot identify all of the α_t and λ_k so we impose the normalizations $\alpha_1 = 1$ and $\lambda_1 = 1$. We also impose the normalization $\mu_1 = 1$. Thus we are left with 109 unknown parameters to estimate.

As starting coefficient values for Model 10, we used the final coefficient estimates from Model 9 plus the starting values $\tau = 0$ and $\rho = 0$. The final log likelihood for Model 10 was 531.13 points higher than the final log likelihood for Model 9 for adding 2 subway time parameters. The resulting R^2 for Model 10 was 0.8383 (the Model 9 R^2 was 0.8256) and the estimated geometric depreciation rate was $\delta^* = 4.52\%$ (21.7). The estimated ϕ was $\phi^* = 1.137$ (104.1). Thus a property without a structure sold at an 13.7% premium compared to a similar property without a structure. As usual, there were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. Also as usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The sequence of relative marginal valuations of land (the λ_k^*) were as follows: $\lambda_0^* = 1.481$ (42.1), $\lambda_1^* = 1$ (imposed restriction), $\lambda_2^* = 1.137$ (35.1), $\lambda_3^* = 1.216$ (33.5) and $\lambda_4^* = 0.930$ (13.3). The sequence of relative marginal valuations of floor space area (the μ_m^*) were as follows: $\mu_0^* = 1.418$ (23.0), $\mu_1^* = 1$ (imposed restriction), $\mu_2^* = 2.255$ (19.3) and $\mu_3^* = 1.540$ (10.2). Thus the estimated λ_k^* and μ_m^* did not change significantly from the estimates for Model 9. The estimated subway time parameter were $\tau^* = -0.0123$ (33.3) and $\rho^* = -0.00606$ (19.1). Thus a 1 minute increase in walking time to the nearest subway station decreases the per square meter land price by 1.2%. A 1 minute time in commuting time to the Tokyo central station decreases land value by 0.6%. The sequence of estimated α_t^* is our land price index for Model 10, P_{L10t}, and this series is plotted in Chart 3 below and is listed in Table A3 of the Appendix.

In our next model, we introduce the number of bedrooms NB_{tn} as a property characteristic that can affect structure value if the property n in quarter t has a structure on it. For the properties in our sample, the number of bedrooms ranged from 2 to 8.[48] Since there were relatively few observations with 6, 7 or 8 bedrooms, we grouped these

[48] For the 5580 properties in our sample that had a structure, the number of observations that had 2, 3,....,8 bedrooms was: 247, 1628, 2441, 841, 295, 90 and 38.

last 3 categories into a single category. Define the bedroom dummy variables $D_{NB,tn,i}$ for observation tn as follows for i = 2,3,4,5; t = 1,...,44 and n = 1,...,N(t):

(35) $D_{NB,tn,i} \equiv 1$ if observation tn has a structure on it with i bedrooms;
$\equiv 0$ elsewhere.

For bedroom group 6, define $D_{NB,tn,6} \equiv 1$ if observation tn has a structure on it with 6, 7 or 8 bedrooms and define $D_{NB,tn,6} \equiv 0$ elsewhere.

Model 11 is the following nonlinear regression:

$$(36)\ V_{tn} = \alpha_t [D_{S,tn} + \phi D_{L,tn}] g_7(x_{tn},y_{tn},\gamma) f_L(L_{tn},\lambda)[1+\tau(TW_{tn}-1)][1+\rho(TT_{tn}-8)] + P_{St}(1-\delta)^{A(t,n)} f_S(S_{tn},\mu)[\Sigma_{i=2}^{6} \kappa_i D_{NB,tn,i}] + \varepsilon_{tn}; \quad t = 1,...,44;\ n = 1,...,N(t).$$

where the all of the functions and parameters which appear in (36) were defined in the previous model except that we have now added 5 bedroom variables, κ_2, κ_3, κ_4, κ_5 and κ_6. We make the same normalizations as we made in Model 10 and in addition, we set $\kappa_2 = 1$. Thus we have added 4 additional unknown κ_i parameters to Model 10 so Model 11 has a total 113 unknown parameters. One might expect the κ_i parameters to monotonically increase as i increases; i.e., more bedrooms indicates a higher quality structure. But we have already introduced the μ_m into our hedonic regression model which allows structure quality to increase as the floor space increases. The correlation between the number of bedrooms and the structure area is 0.93500 and thus there will be a multicollinearity problem in using both of these variables in our nonlinear regression. Thus we cannot expect the κ_i and μ_m to increase monotonically as i and m increase.

As starting coefficient values for Model 11, we used the final coefficient estimates from Model 10 plus the starting values $\kappa_i = 1$ for i = 3,4,5,6.The final log likelihood for Model 11 was 75.03 points higher than the final log likelihood for Model 10 for adding 4 number of bedroom parameters. The resulting R^2 for Model 11 was 0.8400 (the Model 10 R^2 was 0.8383). As usual, there were 3 negative estimates for the land price height parameters: γ_{01}^*, γ_{67}^* and γ_{77}^*. Also as usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. Recall that κ_2 was set equal to 1. The remaining bedroom parameter estimates were as follows: $\kappa_2^* = 1$ (imposed restriction), $\kappa_3^* = 1.1587$ (20.6), $\kappa_4^* = 1.0863$ (21.4), $\kappa_5^* = 0.9116$ (20.0) and $\kappa_6^* = 0.7701$ (18.6). Evidently, a house with more bedrooms did not seem to increase the quality of the structure. The sequence of relative marginal valuations of floor space area (the μ_m^*) were as follows: $\mu_0^* = 1.1573$ (11.1), $\mu_1^* = 1$ (imposed restriction), $\mu_2^* = 2.4201$ (14.2) and $\mu_3^* = 1.7643$ (10.7). Thus, for the most part, houses with more floor space were of higher quality. The collinearity of the number of bedrooms with the floor space of the structure explains the counterintuitive results for the κ_i^*. The remaining parameters for Model 11 were much the same as the Model 10 estimated parameters. As usual, the sequence of estimated α_t^* is our land price series for Model 11, P_{L11t}, and this series is plotted in Chart 3 below and is listed in Table A3 of the Appendix.

The final additional variable that we introduced into our property nonlinear regression model was the width of the land plot, W_{tn} for property sale n in period t. Recall that W_{tn} is measured in 10ths of a meter and the range of this property width variable was 25 to 90. Other residential property hedonic regression models have shown that this variable is a very significant one: the greater is the lot width, the more valuable is the land plot. *Model 12* is the following nonlinear regression:

$$(37) V_{tn} = \alpha_t [D_{S,tn} + \phi D_{L,tn}] g_7(x_{tn},y_{tn},\gamma) f_L(L_{tn},\lambda)[1+\tau(TW_{tn}-1)][1+\rho(TT_{tn}-8)][1+\sigma(W_{tn}-25)] + P_{St}(1-\delta)^{A(t,n)} f_S(S_{tn},\mu)[\Sigma_{i=2}^{6} \kappa_i D_{NB,tn,i}] + \varepsilon_{tn}; \quad t = 1,...,44; n = 1,...,N(t).$$

where the all of the functions and parameters which appear in (37) were defined in the previous model except that we have now added the property width parameter σ. We make the same normalizations as we made in Model 11. Thus we have added 1 additional unknown parameter to Model 11 so Model 12 has a total 114 unknown parameters. Our expectation is that σ will be positive.

As starting coefficient values for Model 12, we used the final coefficient estimates from Model 11 plus the starting value $\sigma = 0$. The final log likelihood for Model 12 was 401.54 points higher than the final log likelihood for Model 11 for adding 1 lot width parameter. The resulting R^2 for Model 12 was 0.8488 (the Model 11 R^2 was 0.8400). The estimate for the lot width parameter was $\sigma^* = 0.00402$ (27.4). Thus an extra meter of lot width adds about 4% to the per meter squared price of the land plot. As usual, the estimated land price height parameters γ_{01}^*, γ_{67}^* and γ_{77}^* turned out to be negative. However, γ_{52}^* also turned out to be negative. But, as usual, the estimated land prices, $g_7(x_{tn},y_{tn},\gamma^*)$ for t = 1,...,T and n = 1,...,N(t), turned out to be positive for each observation in our sample. The estimated parameters for this model are listed in the Appendix; see Table A4. As usual, the sequence of estimated α_t^* is our land price series for Model 12, P_{L12t}, and this series is plotted on Chart 3 is also listed in Table A3 of the Appendix.

Although the fact that Model 12 generated 4 negative estimated γ_{ij}^* did not lead to any negative predicted prices for land for the properties in our sample, these negative estimates could lead to negative land prices for properties not in our sample. Hence, it may be useful to perform a final regression where we restrict the γ_{ij} to be nonnegative.[49] This can be done by replacing γ_{01}, γ_{67}, γ_{77} and γ_{52} in the function $g_7(x_{tn},y_{tn},\gamma)$ by squares of parameters and then rerunning the model defined by (37). *Model 13* is the resulting model. The final log likelihood for Model 13 was 1.19 points lower than the final log likelihood for Model 12 as a result of imposing nonnegativity on the height parameters γ_{01}, γ_{67}, γ_{77} and γ_{52} by entering these parameters as squares in the function $g_7(x_{tn},y_{tn},\gamma)$. The R^2 for Model 13 was 0.8488, which is the same as the R^2 for Model 12. Thus imposing nonnegativity constraints on the γ_{ij} did not lead to a significant loss of fit. The estimated

[49] Nonparametric methods for the estimation of functions of one variable tend to become unreliable for observations close to the boundaries of the domain of definition of the independent variable because the nonparametric method will tend to fit the error terms near the end points of the sample range; see Diewert and Wales (2006; 118) and the literature cited there. The same problem carries over to the nonparametric estimation of surfaces. Forcing the γ_{ij} to be nonnegative will tend to mitigate this problem.

parameters for this model are listed in the Appendix; see Table A5. As usual, the sequence of estimated α_t^* is our land price series for Model 13, P_{L13t}, and this series is listed in Table A3 of the Appendix. It does not appear on Chart 3 because the P_{L13t} are virtually identical to the P_{L12t}.

Our final model in this section is a Ward dummy variable model that adds more explanatory property characteristics to Model 6 in the previous section. This model was defined by equations (25). *Model 14* is defined by the following nonlinear regression model:

$$(38)\ V_{tn} = \alpha_t [D_{S,tn} + \phi D_{L,tn}][\sum_{j=1}^{23} \omega_j D_{W,tn,j}] f_L(L_{tn},\lambda)[1+\tau(TW_{tn}-1)][1+\rho(TT_{tn}-8)][1+\sigma(W_{tn}-25)] + P_{St}(1-\delta)^{A(t,n)} f_S(S_{tn},\mu)[\Sigma_{i=2}^{6} \kappa_i D_{NB,tn,i}] + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

Thus Model 14 is basically the same as Model 12 and 13 except that the Ward dummy variable terms, $\sum_{j=1}^{23} \omega_j D_{W,tn,j}$, replace the Colwell locational grid function, $g_7(x_{tn},y_{tn},\gamma)$, for each observation tn. There are 82 parameters in this model. The final log likelihood for Model 14 was −22492.99, a gain of 478.6 over the final log likelihood of Model 6. However, the log likelihood of Model 14, the Ward dummy variable model with extra characteristics, was 827.00 points below the final log likelihood of Model 13, our "best" Colwell spatial coordinate model. The R^2 for Model 14 was 0.8300 which is not that far below the R^2 for Model 13, which was 0.8488. The estimated parameters for Model 14 are listed in Table A6 in the Appendix.[50] The sequence of land price indexes is the series of estimated coefficients, the α_t^*. This series is labeled as P_{L14t} and is listed in Table A3 and plotted in Chart 3 below.

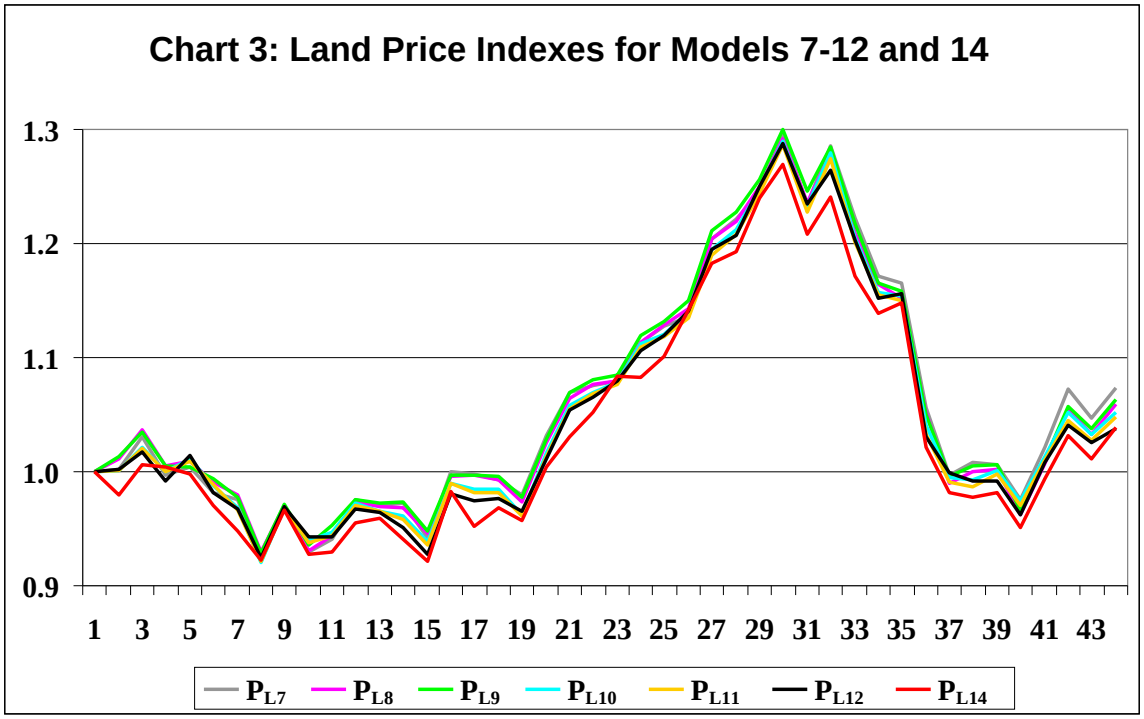
Excluding the location parameters (the γ_{ij}^* for Model 13 and the ω_j^* for Model 14), it can be seen that the remaining parameters (ϕ^*, the λ_k^*, τ^*, ρ^*, σ^*, the μ_m^*, δ^* and the κ_i^*) are roughly similar across Models 13 and 14 and the land price coefficients (the α_t^*) are very close to each other. Our tentative conclusion at this point is that neighbourhood dummy variable models do not fit the data quite as well as a spatial coordinate model but the two types of model generate much the same land prices and hence overall residential property price indexes.[51] Looking at Chart 3, it can be seen that Model 14, the model that used Ward dummy variables to take into account location effects on the price of land, produced the lowest measure of residential land price inflation in Tokyo. Our best spatial coordinate models, Models 12 and 13,[52] had the next lowest measure of land price inflation. The land price indexes generated by Models 7-11 are marginally above the Model 13 and 14 indexes.

---

[50] The estimated parameter values for ϕ, τ, ρ, σ and δ for Models 13 and 14 were 1.125, −0.0130, −0.0668, 0.00402, 0.0417 and 1.155, −0.0136, −0.00945, 0.004393, 0.0402 respectively. Thus the estimates were broadly similar for our best spatial coordinates model and our best Ward dummy variable model.

[51] Hill and Scholz (2018) came to the same conclusion for Sydney overall residential property price indexes.

[52] We did not plot the land price index for Model 13 since it could not be distinguished from the Model 12 index.



In the following section, we compute the overall residential property price indexes that are generated by Models 7-14 and we compare the resulting indexes with a traditional log price time dummy property price index.

8. Overall Residential Property Price Indexes

Models 7-14 in the previous section all have the same general structure in that property value is decomposed into the sum of land value plus structure value plus an error term. For example, using Model 8, the predicted value of property n in quarter t, V_{tn}, is equal to the predicted land value, $\alpha_t^* (D_{S,tn} + \phi^* D_{L,tn}) g_7(x_{tn},y_{tn},\gamma^*) f_L(L_{tn},\lambda^*) \equiv V_{Ltn}$, plus predicted structure value, $P_{St}(1 - \delta^*)^{A(t,n)} S_{tn} \equiv V_{Stn}$. Thus quarter t total predicted land value is $V_{Lt} \equiv \Sigma_{n=1}^{N(t)} V_{Ltn}$ and quarter t total predicted structure value is $V_{St} \equiv \Sigma_{n=1}^{N(t)} V_{Stn}$. The period t price of land for Models 7-14, P_{Lt}, is always α_t^* and the corresponding period t price of a structure is always $p_{St} \equiv P_{St}/P_{S1}$ for t = 1,...,44 where P_{St} is the official structure cost per m^2 of structure. For all models, define the corresponding period t aggregate quantity of land and structure as $Q_{Lt} \equiv V_{Lt}/P_{Lt}$ and $Q_{St} \equiv V_{St}/p_{St}$ for t = 1,...,44. Thus the basic price and quantity data for each model are $(P_{Lt},p_{St},Q_{Lt},Q_{St})$ for t = 1,...,44. The overall property price indexes for Models 7-14 are calculated as Fisher (1922) chained indexes using the price and quantity data for land and structures that has just been defined. Label the resulting *overall property price indexes* for quarter t as P_{7t}, P_{8t}, P_{9t}, P_{10t}, P_{11t}, P_{12t}, P_{13t}, and P_{14t}. These series are listed in Table A7 in the Appendix. As was the case with the corresponding land price indexes, there these overall property price indexes approximate each other fairly closely.

There is one additional overall property price index that we calculate in this section and that is an index that is based on a "traditional" hedonic property price regression that uses the logarithm of price as the dependent variable and has time dummy variables.[53] Define the kth time dummy variable $D_{T,tn,k}$ for property n sold in period t as follows: for t = 1,...,44; n = 1,...,N(t); k = 2,3,...,44:

(39) $D_{T,tn,k} \equiv 1$ if t = k; $D_{T,tn,k} \equiv 0$ if $t \neq k$.

Our best time dummy variable hedonic regression model[54] is the following *Model 15*:

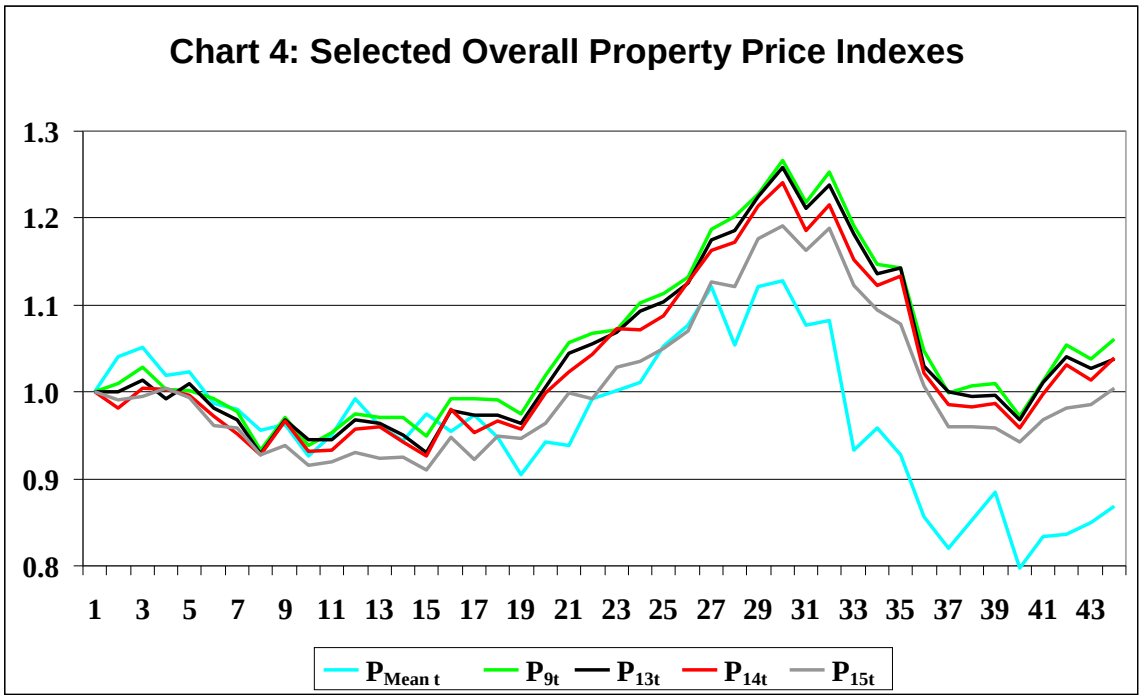
$$(40)\ \ln V_{tn} = \Sigma_{k=2}^{44} \alpha_k D_{T,tn,k} + \sum_{j=1}^{23} \omega_j D_{W,tn,j} + \lambda \ln L_{tn} + \mu S_{tn} + \delta A_{tn} + \tau TW_{tn} + \rho TT_{tn} + \sigma W_{tn} + \Sigma_{i=3}^{6} \kappa_i D_{NB,tn,i} + \varepsilon_{tn}\ ; \qquad t = 1,...,44;\ n = 1,...,N(t).$$

where $\ln V_{tn}$ and $\ln L_{tn}$ denote the natural logarithms of property value V_{tn} and property lot size L_{tn} respectively, the $D_{T,tn,k}$ are time dummy variables, the $D_{W,tn,j}$ are Ward dummy variables, S_{tn} is the floor space area of the property (if there was no structure on the property n in period t, $S_{tn} \equiv 0$), TW_{tn} and TT_{tn} are the subway time variables, W_{tn} is the lot width variable, A_{tn} is the age of the structure on property n sold in period t ($A_{tn} \equiv 0$ if the property had no structure) and the $D_{NB,tn,i}$ are the bedroom dummy variables. The log likelihood of this model cannot be compared to the log likelihood of the previous models because the dependent variable is now the logarithm of the property price instead of the property price. There are 75 unknown parameters in the model defined by equations (40). The R^2 for Model 15 was 0.8323. Set $\alpha_1^* = 0$ and denote the estimated α_2 to α_{44} by α_2^*, α_3^*,..., α_{44}^*. The sequence of overall property price indexes P_{15t} generated by this model are the exponentials of the α_t^*; i.e., define $P_{15t} \equiv \exp[\alpha_t^*]$ for t = 1,...,44. This series is also listed in Table A7 of the Appendix.

Chart 4 below compares several of the overall residential property prices that are defined above: the mean property price index $P_{Mean\ t}$ that appeared in Chart 1 above, P_{9t} (this is based on Model 9 which did not use information on the subway variables, the number of bedrooms and the lot width variable), Model 13 (P_{13t} :our best Colwell spatial coordinates model), Model 14 (P_{14t} :our best Ward dummy variable model) and Model 15 (P_{15t} :our best traditional log price time dummy hedonic regression model that used all of our property price characteristics except the spatial coordinates).

[53] This type of model does not generate reasonable separate land and structure subindexes; see Diewert, Huang and Burnett-Isaacs (2017; 24-25) for an explanation of this assertion.

[54] We ran an initial linear regression using L_{tn} as an independent variable in place of $\ln L_{tn}$. However, this regression had a log likelihood which was 204.99 points lower than our final linear regression defined by (40). The R^2 for this preliminary regression was 0.8274. Note that we could not use $\ln S_{tn}$ as an independent variable because many observations had no structure on them and hence S_{tn} is equal to 0 for these properties and thus we could not take the logarithm of 0.



Several points emerge from a study of Chart 4:

- The mean index, $P_{Mean\ t}$, has a large downward bias as compared to the other 4 indexes which is due to its neglect of structure depreciation. However, the movements in this index are similar to the movements in the other indexes.
- The property price index P_{15t} generated by a traditional log price time dummy hedonic regression model has a downward bias but it is not large.[55]
- The Model 9 property price index, a Colwell spatial coordinates model that used only the 4 fundamental characteristics of a residential property (land plot area, structure floor space area, the age of the structure and some locational variable)[56] generated an overall property price index P_{9t} that is quite close to our best Colwell spatial model, Model 14 which generated the overall property price index P_{14t}. Thus it is probably not necessary for national statistical agencies to collect a great deal of information on housing characteristics in order to produce a decent overall property price index (as well as decent land and structure subindexes).
- The Model 14 property price index, P_{14t}, that used local neighbourhood information about properties instead of spatial coordinate information turned out to be fairly close to our best Colwell spatial index, P_{13t}. Thus following the advice of Hill and Scholz (2018), it is probably not necessary to utilize spatial coordinate information in order to construct a satisfactory overall residential property price index.

[55] Diewert (2010) also observed a similar result.
[56] In addition to these four fundamental variables, we need an exogenous building cost measure in order to implement our basic models.

9. Conclusion

Here are the main points that emerge from our paper:

- Satisfactory residential land price indexes and overall residential property price indexes can be constructed using local neighbourhood dummy variables as explanatory variables in residential property regression models. It is not necessary to use spatial coordinates to model location effects on property prices.
- However, the use of spatial coordinates to model location effects does lead to better fitting regression models.
- The most important housing characteristics information that is needed in order to construct satisfactory residential land and overall property price indexes is information on lot size, floor space area of the property structure (if there is a structure on the property), the age of the structure and some information on the location of the property. In order to obtain a satisfactory land price index, our method requires the use of exogenous information on residential construction costs.
- However, additional information on the characteristics of the property will improve the fit of our hedonic regressions but the effects of the additional information on the resulting land and structure price indexes was minimal for our application to Tokyo residential property price indexes.
- Having land only sales of residential properties should help improve the accuracy of the land price index that is generated by a property regression model. However, for our Japanese data, we found that the value of the land component of a land only property earned a 10-15% premium over the land value of a neighbouring property of the same size but with a structure on the property. We attribute this premium to the costs of demolishing an older structure.
- Our models that used spatial coordinates to account for locational effects on the value of land used Colwell's nonparametric method for fitting a surface. This nonparametric method is much easier to implement than the penalized least squares approach used by Hill and Scholz (2018) to model locational effects on property prices. In section 4 of the paper, we pointed out some of the theoretical advantages of Colwell's method.
- The potential bias in using property price indexes that are based on taking mean or median averages of property prices in a period can be very large. Typically, these methods will have a downward bias due to their neglect of structure depreciation.
- A traditional log price time dummy hedonic regression model that has structure age as an explanatory variable will typically reduce the bias that is inherent in an index based on taking averages of property prices. For our Tokyo data, we found that the traditional hedonic regression model led to an index which had a small downward bias; see Chart 4 in the previous section.

It should be noted that if a national statistical agency were to apply the regression models that were explained in this paper, they would not just run a regression using the entire sample data. A rolling window approach would be used: a window length of say 12 to 16

quarters would be chosen and as the data for each new quarter was processed, the movements in the index over the last two quarters in the sample would be used to update the last published index value; see Shimizu, Nishimura and Watanabe (2010) for an application of this rolling window approach.[57]

Our emphasis in this paper (and in other papers[58]) has been to develop reliable methods for the construction of the land component of residential property price indexes. This task is important for national statistical agencies because the Balance Sheet Accounts in the System of National Accounts requires estimates for the price and volume of land used in production and consumption. In particular, this information is required in order to obtain more accurate estimates of national (and sectoral) Total Factor Productivity growth[59] but for the vast majority of countries, this information is simply not available. We hope that the methods explained in the present paper will be of use to national statistical agencies in developing improved estimates for the price and volume of land used in production and consumption.

Appendix: Supplementary Tables

Table A1: Mean Property Price Index and Model 1 Overall Property Price Index with Land and Structure Subindexes

t	$P_{Mean\ t}$	P_{1t}	P_{L1t}	p_{St}
1	1.00000	1.00000	1.00000	1.00000
2	1.04097	0.98918	0.98918	0.98919
3	1.05097	1.02235	1.02496	0.98919
4	1.01849	0.99183	0.99148	0.99459
5	1.02284	0.98759	0.98779	0.98378
6	0.98708	0.97245	0.97120	0.98919
7	0.97987	0.96833	0.96716	0.98378
8	0.95594	0.91618	0.91246	0.97297
9	0.96213	0.96526	0.96492	0.96216
10	0.92594	0.92538	0.92228	0.96757
11	0.95096	0.93889	0.93699	0.96216
12	0.99168	0.97632	0.97644	0.96757
13	0.96137	0.97547	0.97625	0.95676
14	0.94352	0.97167	0.97257	0.95135
15	0.97532	0.93979	0.93788	0.95676
16	0.95470	0.99830	1.00062	0.95135
17	0.97323	0.98219	0.98344	0.95135
18	0.94821	0.98386	0.98561	0.94595
19	0.90469	0.95997	0.95983	0.95135
20	0.94306	1.02548	1.02965	0.94054
21	0.93779	1.04702	1.05240	0.94054

[57] See Diewert and Fox (2017) for a discussion of other possible methods that could be used to link the indexes generated by successive windows.
[58] See Diewert and Shimizu (2015a) (2015b) (2017a) (2017b) (2019) and Diewert, Fox and Shimizu (2016).
[59] See Jorgenson and Griliches (1967) (1972) who developed the methodology used by national and international statistical agencies to measure TFP growth or Multifactor Productivity growth.

22	**0.99237**	**1.06959**	**1.07597**	**0.94595**
23	**1.00223**	**1.06468**	**1.07080**	**0.94595**
24	**1.01111**	**1.09460**	**1.10251**	**0.94054**
25	**1.05310**	**1.09683**	**1.10519**	**0.93514**
26	**1.07737**	**1.11087**	**1.11907**	**0.95135**
27	**1.12105**	**1.18040**	**1.19230**	**0.95135**
28	**1.05456**	**1.19449**	**1.20716**	**0.95135**
29	**1.12150**	**1.21903**	**1.23350**	**0.94595**
30	**1.12764**	**1.26251**	**1.27970**	**0.94054**
31	**1.07651**	**1.21329**	**1.22695**	**0.95676**
32	**1.08274**	**1.25004**	**1.26592**	**0.95676**
33	**0.93306**	**1.19154**	**1.20469**	**0.94054**
34	**0.95904**	**1.14413**	**1.15314**	**0.95135**
35	**0.92830**	**1.15154**	**1.15982**	**0.96757**
36	**0.85665**	**1.04925**	**1.05014**	**0.96216**
37	**0.82004**	**0.99146**	**0.98891**	**0.95135**
38	**0.85241**	**1.01351**	**1.01210**	**0.95676**
39	**0.88506**	**1.00130**	**0.99793**	**0.97297**
40	**0.79792**	**0.97515**	**0.97133**	**0.95676**
41	**0.83303**	**1.01996**	**1.01836**	**0.95676**
42	**0.83594**	**1.06940**	**1.07088**	**0.95135**
43	**0.85040**	**1.04038**	**1.03972**	**0.96216**
44	**0.86796**	**1.06665**	**1.06734**	**0.96216**

Table A2: Land Price Indexes for Models 1-6

t	**P_{L1t}**	**P_{L2t}**	**P_{L3t}**	**P_{L4t}**	**P_{L5t}**	**P_{L6t}**
1	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**
2	**0.98918**	**0.99562**	**0.99083**	**1.00571**	**0.99755**	**0.99373**
3	**1.02496**	**1.02431**	**1.02241**	**1.03145**	**1.03206**	**1.02754**
4	**0.99148**	**0.99702**	**0.99306**	**0.99861**	**0.99921**	**1.01174**
5	**0.98779**	**0.99087**	**0.98804**	**1.00480**	**0.99546**	**0.99992**
6	**0.97120**	**0.98053**	**0.97689**	**0.98820**	**0.98081**	**0.98339**
7	**0.96716**	**0.97469**	**0.97183**	**0.98043**	**0.97361**	**0.96590**
8	**0.91246**	**0.91578**	**0.91658**	**0.92612**	**0.91483**	**0.93295**
9	**0.96492**	**0.96090**	**0.96328**	**0.97122**	**0.96323**	**0.97884**
10	**0.92228**	**0.91627**	**0.91697**	**0.93082**	**0.92485**	**0.92142**
11	**0.93699**	**0.92883**	**0.93167**	**0.93960**	**0.92569**	**0.93043**
12	**0.97644**	**0.97541**	**0.97671**	**0.97791**	**0.97285**	**0.96467**
13	**0.97625**	**0.97251**	**0.97435**	**0.97614**	**0.97267**	**0.97691**
14	**0.97257**	**0.96853**	**0.97079**	**0.97334**	**0.96921**	**0.96688**
15	**0.93788**	**0.93576**	**0.93248**	**0.93682**	**0.93110**	**0.93360**
16	**1.00062**	**1.00570**	**1.00540**	**1.00531**	**1.00672**	**1.00979**
17	**0.98344**	**0.99016**	**0.98658**	**0.99512**	**0.98592**	**0.97770**
18	**0.98561**	**0.99344**	**0.98815**	**0.99427**	**0.98895**	**0.99493**
19	**0.95983**	**0.96747**	**0.96331**	**0.97820**	**0.96839**	**0.97260**
20	**1.02965**	**1.02773**	**1.02756**	**1.03513**	**1.03001**	**1.03566**
21	**1.05240**	**1.05913**	**1.05877**	**1.07369**	**1.05837**	**1.04751**
22	**1.07597**	**1.07920**	**1.07865**	**1.07899**	**1.07835**	**1.07413**
23	**1.07080**	**1.07246**	**1.07464**	**1.08646**	**1.07802**	**1.10446**
24	**1.10251**	**1.10792**	**1.10687**	**1.11603**	**1.11026**	**1.09147**
25	**1.10519**	**1.11619**	**1.10947**	**1.12826**	**1.11054**	**1.11076**
26	**1.11907**	**1.13244**	**1.12064**	**1.13779**	**1.12835**	**1.14190**

27	**1.19230**	**1.19785**	**1.19483**	**1.20930**	**1.19765**	**1.20595**
28	**1.20716**	**1.20434**	**1.19851**	**1.22033**	**1.20889**	**1.20877**
29	**1.23350**	**1.23573**	**1.22889**	**1.24735**	**1.23748**	**1.24309**
30	**1.27970**	**1.28027**	**1.28135**	**1.29594**	**1.28089**	**1.28484**
31	**1.22695**	**1.23436**	**1.22913**	**1.23777**	**1.23266**	**1.22324**
32	**1.26592**	**1.26387**	**1.26272**	**1.28812**	**1.27056**	**1.26938**
33	**1.20469**	**1.21042**	**1.21491**	**1.22545**	**1.21578**	**1.18474**
34	**1.15314**	**1.15323**	**1.15710**	**1.17238**	**1.15621**	**1.17714**
35	**1.15982**	**1.15548**	**1.15607**	**1.16891**	**1.16001**	**1.17180**
36	**1.05014**	**1.05138**	**1.05357**	**1.05976**	**1.04599**	**1.05754**
37	**0.98891**	**0.98995**	**0.99088**	**1.00107**	**0.99035**	**0.98920**
38	**1.01210**	**1.00258**	**1.00928**	**1.01280**	**1.00928**	**1.00939**
39	**0.99793**	**1.00305**	**1.00200**	**1.01316**	**1.00827**	**1.00484**
40	**0.97133**	**0.96964**	**0.97313**	**0.98484**	**0.97440**	**0.97225**
41	**1.01836**	**1.01128**	**1.01712**	**1.03069**	**1.01728**	**1.02594**
42	**1.07088**	**1.07241**	**1.07038**	**1.08017**	**1.07497**	**1.07719**
43	**1.03972**	**1.03946**	**1.04408**	**1.05292**	**1.04155**	**1.04387**
44	**1.06734**	**1.06518**	**1.06341**	**1.07786**	**1.06898**	**1.08631**

Table A3: Land Price Indexes for Models 7-14

t	**P_{L7t}**	**P_{L8t}**	**P_{L9t}**	**P_{L10t}**	**P_{L11t}**	**P_{L12t}**	**P_{L13t}**	**P_{L14t}**
1	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**	**1.00000**
2	**1.00245**	**1.01090**	**1.01292**	**1.00101**	**1.00111**	**1.00164**	**1.00145**	**0.97977**
3	**1.03037**	**1.03702**	**1.03487**	**1.02124**	**1.02007**	**1.01684**	**1.01688**	**1.00586**
4	**0.99704**	**1.00554**	**1.00462**	**1.00023**	**0.99957**	**0.99137**	**0.99141**	**1.00403**
5	**1.00434**	**1.00909**	**1.00382**	**1.01092**	**1.00968**	**1.01428**	**1.01405**	**0.99830**
6	**0.98205**	**0.99069**	**0.99359**	**0.98811**	**0.98858**	**0.98148**	**0.98140**	**0.97022**
7	**0.97562**	**0.97917**	**0.97709**	**0.96830**	**0.96677**	**0.96747**	**0.96752**	**0.94775**
8	**0.92170**	**0.92926**	**0.92907**	**0.92076**	**0.92111**	**0.92465**	**0.92460**	**0.92279**
9	**0.96832**	**0.97006**	**0.97187**	**0.96830**	**0.96860**	**0.96908**	**0.96924**	**0.96664**
10	**0.93002**	**0.93092**	**0.93577**	**0.93984**	**0.93767**	**0.94248**	**0.94245**	**0.92721**
11	**0.94091**	**0.94415**	**0.95293**	**0.94658**	**0.94352**	**0.94314**	**0.94278**	**0.92937**
12	**0.97543**	**0.97350**	**0.97551**	**0.97221**	**0.97030**	**0.96741**	**0.96761**	**0.95536**
13	**0.97048**	**0.96972**	**0.97205**	**0.96535**	**0.96488**	**0.96458**	**0.96464**	**0.95959**
14	**0.97196**	**0.96837**	**0.97328**	**0.96139**	**0.95836**	**0.95101**	**0.95112**	**0.94111**
15	**0.94163**	**0.94614**	**0.94808**	**0.93870**	**0.93570**	**0.92720**	**0.92735**	**0.92104**
16	**1.00029**	**0.99578**	**0.99728**	**0.98952**	**0.98997**	**0.98109**	**0.98122**	**0.98288**
17	**0.99845**	**0.99687**	**0.99707**	**0.98445**	**0.98197**	**0.97465**	**0.97472**	**0.95207**
18	**0.99277**	**0.99321**	**0.99575**	**0.98498**	**0.98180**	**0.97621**	**0.97614**	**0.96866**
19	**0.97953**	**0.97349**	**0.97763**	**0.96176**	**0.96150**	**0.96500**	**0.96504**	**0.95761**
20	**1.03172**	**1.02541**	**1.02813**	**1.01653**	**1.01355**	**1.01088**	**1.01092**	**1.00430**
21	**1.06968**	**1.06459**	**1.06959**	**1.05843**	**1.05531**	**1.05412**	**1.05419**	**1.03065**
22	**1.07515**	**1.07634**	**1.08072**	**1.06978**	**1.06841**	**1.06518**	**1.06534**	**1.05244**
23	**1.07988**	**1.07917**	**1.08454**	**1.07874**	**1.07681**	**1.08001**	**1.08012**	**1.08375**
24	**1.11303**	**1.11331**	**1.11888**	**1.11199**	**1.10909**	**1.10623**	**1.10614**	**1.08280**
25	**1.12746**	**1.12879**	**1.13146**	**1.12075**	**1.11876**	**1.11938**	**1.11936**	**1.10126**
26	**1.13621**	**1.14314**	**1.15034**	**1.13941**	**1.13519**	**1.14080**	**1.14095**	**1.14218**
27	**1.20393**	**1.20416**	**1.21079**	**1.19467**	**1.18940**	**1.19447**	**1.19453**	**1.18262**
28	**1.22165**	**1.21905**	**1.22756**	**1.21229**	**1.20788**	**1.20719**	**1.20730**	**1.19269**

29	**1.24442**	**1.25027**	**1.25641**	**1.24786**	**1.24431**	**1.24997**	**1.24994**	**1.24013**
30	**1.29517**	**1.29206**	**1.30048**	**1.28978**	**1.28683**	**1.28815**	**1.28809**	**1.26915**
31	**1.23490**	**1.23590**	**1.24606**	**1.23307**	**1.22783**	**1.23456**	**1.23473**	**1.20812**
32	**1.28560**	**1.27975**	**1.28514**	**1.27937**	**1.27498**	**1.26404**	**1.26402**	**1.24084**
33	**1.22290**	**1.20928**	**1.21732**	**1.20556**	**1.20131**	**1.20324**	**1.20334**	**1.17161**
34	**1.17110**	**1.16392**	**1.16572**	**1.15735**	**1.15470**	**1.15226**	**1.15209**	**1.13840**
35	**1.16527**	**1.15226**	**1.15855**	**1.15415**	**1.14983**	**1.15653**	**1.15666**	**1.14788**
36	**1.05617**	**1.05242**	**1.05050**	**1.04062**	**1.03220**	**1.03011**	**1.03001**	**1.02186**
37	**0.99660**	**0.98956**	**0.99613**	**0.99526**	**0.99077**	**0.99864**	**0.99856**	**0.98180**
38	**1.00818**	**1.00007**	**1.00491**	**0.99335**	**0.98698**	**0.99171**	**0.99186**	**0.97776**
39	**1.00612**	**1.00231**	**1.00577**	**1.00131**	**0.99751**	**0.99182**	**0.99191**	**0.98129**
40	**0.97594**	**0.96214**	**0.96592**	**0.97429**	**0.97012**	**0.96264**	**0.96261**	**0.95144**
41	**1.02111**	**1.00718**	**1.01057**	**1.01328**	**1.01020**	**1.00866**	**1.00860**	**0.99438**
42	**1.07266**	**1.05556**	**1.05708**	**1.05184**	**1.04489**	**1.04113**	**1.04124**	**1.03131**
43	**1.04658**	**1.03255**	**1.03780**	**1.03271**	**1.02745**	**1.02551**	**1.02553**	**1.01140**
44	**1.07361**	**1.05879**	**1.06364**	**1.05217**	**1.04833**	**1.03735**	**1.03739**	**1.03911**

Table A4: Model 12 Estimated Parameters

Coef	Estimate	T stat	Coef	Estimate	T stat	Coef	Estimate	T stat
ϕ	**1.1250**	**108.12**	γ_{55}^*	**2.0007**	**11.40**	α_{22}^*	**1.0652**	**70.67**
γ_{01}^*	**−178.4700**	**−1.20**	γ_{65}^*	**2.5393**	**13.87**	α_{23}^*	**1.0800**	**71.14**
γ_{11}^*	**−0.4677**	**−0.88**	γ_{75}^*	**3.0373**	**8.29**	α_{24}^*	**1.1062**	**72.99**
γ_{20}^*	**2.5925**	**2.90**	γ_{06}^*	**2.4371**	**10.57**	α_{25}^*	**1.1194**	**72.94**
γ_{21}^*	**5.3420**	**37.90**	γ_{16}^*	**3.6399**	**22.20**	α_{26}^*	**1.1408**	**74.26**
γ_{30}^*	**3.9269**	**13.32**	γ_{26}^*	**3.3394**	**25.68**	α_{27}^*	**1.1945**	**74.79**
γ_{31}^*	**3.7477**	**34.44**	γ_{36}^*	**2.6204**	**17.32**	α_{28}^*	**1.2072**	**75.94**
γ_{40}^*	**3.7436**	**4.59**	γ_{46}^*	**1.8189**	**9.70**	α_{29}^*	**1.2500**	**75.85**
γ_{41}^*	**2.7082**	**6.93**	γ_{56}^*	**2.5317**	**16.33**	α_{30}^*	**1.2882**	**75.82**
γ_{02}^*	**3.5099**	**6.44**	γ_{66}^*	**2.9012**	**12.77**	α_{31}^*	**1.2346**	**71.35**
γ_{12}^*	**4.8376**	**35.22**	γ_{76}^*	**1.1706**	**1.05**	α_{32}^*	**1.2640**	**72.84**
γ_{22}^*	**6.2825**	**42.02**	γ_{17}^*	**0.8890**	**0.42**	α_{33}^*	**1.2032**	**66.80**
γ_{32}^*	**4.4467**	**37.70**	γ_{27}^*	**2.2367**	**2.56**	α_{34}^*	**1.1523**	**65.75**
γ_{42}^*	**8.6903**	**29.12**	γ_{37}^*	**6.1267**	**3.57**	α_{35}^*	**1.1565**	**62.33**
γ_{52}^*	**-1.3545**	**-0.12**	γ_{47}^*	**2.1635**	**6.72**	α_{36}^*	**1.0301**	**52.86**
γ_{62}^*	**3.0835**	**0.58**	γ_{57}^*	**1.8749**	**4.88**	α_{37}^*	**0.9986**	**50.28**
γ_{72}^*	**11.1680**	**0.44**	γ_{67}^*	**-0.7877**	**-0.71**	α_{38}^*	**0.9917**	**53.61**
γ_{03}^*	**3.7105**	**11.70**	γ_{77}^*	**4.4102**	**0.17**	α_{39}^*	**0.9918**	**62.16**
γ_{13}^*	**4.6941**	**35.58**	α_2^*	**1.0016**	**68.99**	α_{40}^*	**0.9626**	**57.74**
γ_{23}^*	**5.2365**	**41.73**	α_3^*	**1.0168**	**62.78**	α_{41}^*	**1.0087**	**61.79**
γ_{33}^*	**9.0513**	**41.68**	α_4^*	**0.9914**	**65.12**	α_{42}^*	**1.0411**	**63.60**
γ_{43}^*	**4.0106**	**12.52**	α_5^*	**1.0143**	**67.08**	α_{43}^*	**1.0255**	**65.01**
γ_{53}^*	**3.5972**	**11.75**	α_6^*	**0.9815**	**67.75**	α_{44}^*	**1.0374**	**65.20**
γ_{63}^*	**3.0966**	**11.06**	α_7^*	**0.9675**	**65.21**	λ_0^*	**1.4970**	**41.09**
γ_{73}^*	**2.5846**	**2.61**	α_8^*	**0.9247**	**68.34**	λ_2^*	**1.1260**	**34.84**
γ_{04}^*	**4.4644**	**23.67**	α_9^*	**0.9691**	**64.08**	λ_3^*	**1.2050**	**33.81**
γ_{14}^*	**4.5370**	**41.16**	α_{10}^*	**0.9425**	**64.13**	λ_4^*	**0.9791**	**14.61**

γ_{24}^*	4.3453	39.08	α_{11}^*	0.9431	60.79	τ^*	-0.0130	-36.33
γ_{34}^*	5.2514	37.19	α_{12}^*	0.9674	68.93	ρ^*	-0.0068	-23.05
γ_{44}^*	7.1714	33.87	α_{13}^*	0.9646	62.57	σ^*	0.0040	27.38
γ_{54}^*	3.4810	18.90	α_{14}^*	0.9510	62.63	μ_0^*	1.1063	10.59
γ_{64}^*	2.6694	14.46	α_{15}^*	0.9272	58.85	μ_2^*	2.3519	13.99
γ_{74}^*	2.7069	10.41	α_{16}^*	0.9811	64.31	μ_3^*	1.6727	10.27
γ_{05}^*	3.3997	28.40	α_{17}^*	0.9747	64.07	δ^*	0.0416	22.63
γ_{15}^*	3.5602	33.64	α_{18}^*	0.9762	67.01	κ_3^*	1.1456	20.14
γ_{25}^*	3.8913	35.09	α_{19}^*	0.9650	64.50	κ_4^*	1.0759	20.97
γ_{35}^*	5.0797	36.50	α_{20}^*	1.0109	70.80	κ_5^*	0.9028	19.55
γ_{45}^*	4.2398	31.73	α_{21}^*	1.0541	70.31	κ_6^*	0.7627	17.94

Table A5: Model 13 Estimated Parameters

Coef	Estimate	T stat	Coef	Estimate	T stat	Coef	Estimate	T stat
ϕ	1.12500	105.1	γ_{55}^*	1.99910	10.5	α_{22}^*	1.06530	69.7
γ_{01}^*	0.00000	n.a.	γ_{65}^*	2.55230	13.1	α_{23}^*	1.08010	70.0
γ_{11}^*	0.00000	n.a.	γ_{75}^*	3.01560	7.6	α_{24}^*	1.10610	72.2
γ_{20}^*	2.61950	3.0	γ_{06}^*	2.43640	9.9	α_{25}^*	1.11940	71.7
γ_{21}^*	5.32700	36.7	γ_{16}^*	3.63810	19.8	α_{26}^*	1.14100	73.2
γ_{30}^*	3.91830	12.9	γ_{26}^*	3.33900	24.0	α_{27}^*	1.19450	72.9
γ_{31}^*	3.75460	32.9	γ_{36}^*	2.62000	16.0	α_{28}^*	1.20730	74.5
γ_{40}^*	3.76010	4.6	γ_{46}^*	1.81930	9.2	α_{29}^*	1.24990	73.8
γ_{41}^*	2.68760	6.7	γ_{56}^*	2.53640	15.3	α_{30}^*	1.28810	73.9
γ_{02}^*	3.48220	5.9	γ_{66}^*	2.85420	12.3	α_{31}^*	1.23470	70.0
γ_{12}^*	4.80540	34.2	γ_{76}^*	1.32220	1.1	α_{32}^*	1.26400	71.5
γ_{22}^*	6.27970	40.4	γ_{17}^*	0.89163	0.4	α_{33}^*	1.20330	66.5
γ_{32}^*	4.44750	36.5	γ_{27}^*	2.23780	2.4	α_{34}^*	1.15210	65.1
γ_{42}^*	8.68450	27.9	γ_{37}^*	6.11950	3.5	α_{35}^*	1.15670	62.8
γ_{52}^*	0.00000	n.a	γ_{47}^*	2.18610	6.4	α_{36}^*	1.03000	54.1
γ_{62}^*	3.04700	0.6	γ_{57}^*	1.80240	4.5	α_{37}^*	0.99856	51.3
γ_{72}^*	11.28900	0.4	γ_{67}^*	0.00000	n.a.	α_{38}^*	0.99186	54.8
γ_{03}^*	3.71350	11.1	γ_{77}^*	2.37400	0.1	α_{39}^*	0.99191	62.1
γ_{13}^*	4.70260	33.5	α_2^*	1.00150	68.6	α_{40}^*	0.96261	58.6
γ_{23}^*	5.23550	40.5	α_3^*	1.01690	62.8	α_{41}^*	1.00860	62.1
γ_{33}^*	9.04860	40.2	α_4^*	0.99141	64.9	α_{42}^*	1.04120	63.8
γ_{43}^*	4.01050	12.2	α_5^*	1.01400	66.4	α_{43}^*	1.02550	64.5
γ_{53}^*	3.58740	11.6	α_6^*	0.98140	67.3	α_{44}^*	1.03740	65.1
γ_{63}^*	3.09950	10.8	α_7^*	0.96752	65.5	λ_0^*	1.49740	40.8
γ_{73}^*	2.57700	2.6	α_8^*	0.92460	68.1	λ_2^*	1.12650	35.3
γ_{04}^*	4.46370	22.3	α_9^*	0.96924	64.1	λ_3^*	1.20520	35.6
γ_{14}^*	4.53470	39.5	α_{10}^*	0.94245	64.2	λ_4^*	0.97874	15.3
γ_{24}^*	4.34410	37.4	α_{11}^*	0.94278	61.1	τ^*	–0.01303	–39.6
γ_{34}^*	5.25070	35.7	α_{12}^*	0.96761	68.8	ρ^*	–0.00684	–22.7

γ_{44}^*	7.16970	32.0	α_{13}^*	0.96464	62.7	σ^*	0.00402	26.3
γ_{54}^*	3.48320	18.2	α_{14}^*	0.95112	62.6	μ_0^*	1.10630	10.0
γ_{64}^*	2.66450	13.6	α_{15}^*	0.92735	59.2	μ_2^*	2.35450	13.3
γ_{74}^*	2.71140	10.1	α_{16}^*	0.98122	64.0	μ_3^*	1.67270	10.0
γ_{05}^*	3.39910	26.0	α_{17}^*	0.97472	64.1	δ^*	0.04166	18.4
γ_{15}^*	3.56030	31.4	α_{18}^*	0.97614	67.0	κ_3^*	1.14540	18.5
γ_{25}^*	3.89100	32.8	α_{19}^*	0.96504	64.2	κ_4^*	1.07570	19.3
γ_{35}^*	5.07840	34.5	α_{20}^*	1.01090	70.0	κ_5^*	0.90272	17.8
γ_{45}^*	4.23890	30.3	α_{21}^*	1.05420	69.7	κ_6^*	0.76241	16.3

Table A6: Model 14 Estimated Parameters

Coef	Estimate	T stat	Coef	Estimate	T stat	Coef	Estimate	T stat
ϕ	1.15540	96.9	α_{29}^*	1.24010	69.6	ω_{13}^*	7.01990	40.9
α_2^*	0.97977	64.3	α_{30}^*	1.26920	70.2	ω_{14}^*	4.57510	39.1
α_3^*	1.00590	59.6	α_{31}^*	1.20810	66.0	ω_{15}^*	4.82300	41.4
α_4^*	1.00400	61.2	α_{32}^*	1.24080	67.6	ω_{16}^*	4.99440	38.0
α_5^*	0.99830	62.6	α_{33}^*	1.17160	62.6	ω_{17}^*	3.57980	25.8
α_6^*	0.97022	63.8	α_{34}^*	1.13840	61.8	ω_{18}^*	3.12610	21.9
α_7^*	0.94775	61.9	α_{35}^*	1.14790	58.8	ω_{19}^*	3.83910	34.0
α_8^*	0.92279	63.8	α_{36}^*	1.02190	51.7	ω_{20}^*	4.07610	38.2
α_9^*	0.96664	60.6	α_{37}^*	0.98180	48.3	ω_{21}^*	2.46420	27.4
α_{10}^*	0.92721	60.4	α_{38}^*	0.97776	51.9	ω_{22}^*	2.79550	25.2
α_{11}^*	0.92937	58.3	α_{39}^*	0.98129	58.7	ω_{23}^*	2.99060	28.9
α_{12}^*	0.95536	65.0	α_{40}^*	0.95144	55.5	λ_0^*	1.46500	38.7
α_{13}^*	0.95959	58.9	α_{41}^*	0.99438	58.6	λ_2^*	1.11790	33.9
α_{14}^*	0.94111	59.5	α_{42}^*	1.03130	60.3	λ_3^*	1.20540	34.7
α_{15}^*	0.92104	56.1	α_{43}^*	1.01140	60.8	λ_4^*	0.93182	14.3
α_{16}^*	0.98288	60.5	α_{44}^*	1.03910	61.4	τ^*	–0.01363	–40.6
α_{17}^*	0.95207	60.7	ω_1^*	7.77560	11.7	ρ^*	–0.00945	–43.6
α_{18}^*	0.96866	63.6	ω_2^*	5.39740	8.7	σ^*	0.00393	26.2
α_{19}^*	0.95761	60.7	ω_3^*	7.42060	40.1	μ_0^*	1.24560	10.4
α_{20}^*	1.00430	66.2	ω_4^*	5.56940	40.6	μ_2^*	2.57690	13.0
α_{21}^*	1.03070	65.9	ω_5^*	5.69080	41.1	μ_3^*	1.69770	9.4
α_{22}^*	1.05240	65.8	ω_6^*	4.77040	20.5	δ^*	0.04024	18.3
α_{23}^*	1.08380	66.1	ω_7^*	3.14740	21.2	κ_3^*	1.10660	18.5
α_{24}^*	1.08280	68.5	ω_8^*	3.71300	25.5	κ_4^*	1.04140	19.3
α_{25}^*	1.10130	68.0	ω_9^*	5.28570	40.4	κ_5^*	0.85249	18.2
α_{26}^*	1.14220	69.2	ω_{10}^*	6.49860	41.6	κ_6^*	0.73501	17.0
α_{27}^*	1.18260	69.0	ω_{11}^*	4.81450	41.1			
α_{28}^*	1.19270	70.7	ω_{12}^*	5.76630	41.7			

Table A7: Overall Property Price Indexes for Models 7-15

t	P_{7t}	P_{8t}	P_{9t}	P_{10t}	P_{11t}	P_{12t}	P_{13t}	P_{14t}	P_{15t}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

2	**1.00085**	**1.00845**	**1.00949**	**0.99938**	**0.99952**	**1.00004**	**0.99988**	**0.98103**	**0.99032**
3	**1.02558**	**1.03182**	**1.02850**	**1.01703**	**1.01615**	**1.01344**	**1.01347**	**1.00391**	**0.99496**
4	**0.99679**	**1.00434**	**1.00317**	**0.99949**	**0.99896**	**0.99176**	**0.99179**	**1.00301**	**1.00487**
5	**1.00195**	**1.00631**	**1.00097**	**1.00730**	**1.00635**	**1.01043**	**1.01023**	**0.99662**	**0.99311**
6	**0.98255**	**0.99021**	**0.99266**	**0.98780**	**0.98825**	**0.98182**	**0.98175**	**0.97234**	**0.96075**
7	**0.97621**	**0.97920**	**0.97732**	**0.96943**	**0.96805**	**0.96862**	**0.96867**	**0.95153**	**0.95844**
8	**0.92624**	**0.93260**	**0.93333**	**0.92566**	**0.92588**	**0.92890**	**0.92886**	**0.92799**	**0.92797**
9	**0.96803**	**0.96939**	**0.97086**	**0.96774**	**0.96810**	**0.96850**	**0.96865**	**0.96685**	**0.93849**
10	**0.93363**	**0.93392**	**0.93903**	**0.94262**	**0.94067**	**0.94489**	**0.94488**	**0.93178**	**0.91544**
11	**0.94298**	**0.94553**	**0.95376**	**0.94812**	**0.94538**	**0.94495**	**0.94464**	**0.93316**	**0.92018**
12	**0.97477**	**0.97282**	**0.97451**	**0.97169**	**0.96999**	**0.96738**	**0.96756**	**0.95706**	**0.93083**
13	**0.96926**	**0.96843**	**0.97024**	**0.96441**	**0.96401**	**0.96375**	**0.96382**	**0.95974**	**0.92352**
14	**0.97001**	**0.96669**	**0.97068**	**0.96029**	**0.95762**	**0.95107**	**0.95117**	**0.94272**	**0.92530**
15	**0.94357**	**0.94721**	**0.94931**	**0.94094**	**0.93829**	**0.93059**	**0.93073**	**0.92570**	**0.91027**
16	**0.99632**	**0.99216**	**0.99266**	**0.98602**	**0.98654**	**0.97875**	**0.97888**	**0.98075**	**0.94829**
17	**0.99465**	**0.99315**	**0.99247**	**0.98147**	**0.97936**	**0.97294**	**0.97301**	**0.95307**	**0.92278**
18	**0.98899**	**0.98933**	**0.99068**	**0.98135**	**0.97862**	**0.97378**	**0.97372**	**0.96728**	**0.94944**
19	**0.97747**	**0.97174**	**0.97505**	**0.96100**	**0.96086**	**0.96415**	**0.96419**	**0.95787**	**0.94630**
20	**1.02446**	**1.01891**	**1.01975**	**1.00990**	**1.00738**	**1.00521**	**1.00526**	**0.99945**	**0.96381**
21	**1.05930**	**1.05513**	**1.05731**	**1.04807**	**1.04548**	**1.04480**	**1.04487**	**1.02350**	**0.99846**
22	**1.06477**	**1.06639**	**1.06788**	**1.05887**	**1.05789**	**1.05536**	**1.05552**	**1.04381**	**0.99253**
23	**1.06913**	**1.06901**	**1.07136**	**1.06707**	**1.06560**	**1.06902**	**1.06914**	**1.07255**	**1.02813**
24	**1.09931**	**1.10035**	**1.10220**	**1.09710**	**1.09481**	**1.09275**	**1.09268**	**1.07122**	**1.03585**
25	**1.11200**	**1.11415**	**1.11297**	**1.10449**	**1.10305**	**1.10421**	**1.10421**	**1.08744**	**1.04984**
26	**1.12146**	**1.12868**	**1.13160**	**1.12295**	**1.11950**	**1.12518**	**1.12532**	**1.12597**	**1.07056**
27	**1.18379**	**1.18522**	**1.18652**	**1.17342**	**1.16902**	**1.17434**	**1.17440**	**1.16287**	**1.12662**
28	**1.20005**	**1.19898**	**1.20171**	**1.18946**	**1.18586**	**1.18597**	**1.18607**	**1.17204**	**1.12081**
29	**1.22028**	**1.22718**	**1.22703**	**1.22104**	**1.21831**	**1.22430**	**1.22429**	**1.21440**	**1.17653**
30	**1.26637**	**1.26535**	**1.26639**	**1.25869**	**1.25662**	**1.25878**	**1.25873**	**1.24039**	**1.19156**
31	**1.21231**	**1.21459**	**1.21843**	**1.20829**	**1.20404**	**1.21096**	**1.21113**	**1.18599**	**1.16233**
32	**1.25840**	**1.25478**	**1.25352**	**1.25011**	**1.24667**	**1.23771**	**1.23771**	**1.21555**	**1.18834**
33	**1.20014**	**1.18908**	**1.19125**	**1.18219**	**1.17877**	**1.18123**	**1.18133**	**1.15166**	**1.12278**
34	**1.15480**	**1.14904**	**1.14690**	**1.14051**	**1.13833**	**1.13665**	**1.13651**	**1.12321**	**1.09411**
35	**1.15144**	**1.14019**	**1.14265**	**1.13964**	**1.13588**	**1.14229**	**1.14242**	**1.13354**	**1.07889**
36	**1.05406**	**1.05014**	**1.04768**	**1.03922**	**1.03148**	**1.02968**	**1.02960**	**1.02197**	**1.00741**
37	**0.99992**	**0.99258**	**0.99872**	**0.99792**	**0.99355**	**1.00047**	**1.00040**	**0.98529**	**0.95965**
38	**1.01088**	**1.00263**	**1.00714**	**0.99688**	**0.99081**	**0.99490**	**0.99504**	**0.98235**	**0.95935**
39	**1.01078**	**1.00625**	**1.00985**	**1.00584**	**1.00207**	**0.99677**	**0.99685**	**0.98735**	**0.95880**
40	**0.98161**	**0.96775**	**0.97202**	**0.97964**	**0.97553**	**0.96856**	**0.96853**	**0.95858**	**0.94232**
41	**1.02323**	**1.00961**	**1.01282**	**1.01541**	**1.01229**	**1.01091**	**1.01086**	**0.99794**	**0.96764**
42	**1.06985**	**1.05379**	**1.05435**	**1.04994**	**1.04330**	**1.04003**	**1.04014**	**1.03096**	**0.98085**
43	**1.04683**	**1.03331**	**1.03785**	**1.03344**	**1.02837**	**1.02667**	**1.02670**	**1.01379**	**0.98602**
44	**1.07165**	**1.05763**	**1.06135**	**1.05123**	**1.04746**	**1.03754**	**1.03758**	**1.03911**	**1.00378**

References

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

Bizley, M.T.L. (1958), "A Measure of Smoothness and Some Remarks on a New Principle of Graduation", *Journal of the Institute of Actuaries* 84, 125-165.

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

Buja, A., T. Hastie and R. Tibshirani (1989), "Linear Smoothers and Additive Models", *Annals of Statistics* 18, 454-509.

Burnett-Issacs, K., N. Huang and W.E. Diewert (2016), "Developing Land and Structure Price Indexes for Ottawa Condominium Apartments", Discussion Paper 16-09, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada.

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

Colwell, P. F. (1998), "A Primer on Piecewise Parabolic Multiple Regression Analysis via Estimations of Chicago CBD Land Prices", *Journal of Real Estate Finance and Economics* 17:1, 87-97.

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., W.E. Diewert (eds.) (2013), *Residential Property Price Indices Handbook*, Luxembourg: Eurostat.

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. (2008), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Further Directions", pp. 11-36 in *Proceedings From the OECD Workshop on Productivity Measurement and Analysis*, Paris: OECD.

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. and K.J. Fox (2017), "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data, Discussion Paper 17-02, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1L4.

Diewert, W.E., K.J. Fox and C. Shimizu (2016), "Commercial Property Price Indexes and the System of National Accounts", *Journal of Economic Surveys* 30, 913-943.

Diewert, W.E., J. de Haan and R. Hendriks (2011), "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 (2015), "Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components", *Econometric Reviews*, 34:1-2, 106-126.

Diewert, W.E., N. Huang and K. Burnett-Isaacs (2017), "Alternative Approaches for Resale Housing Price Indexes", Discussion Paper 17-05, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1L4.

Diewert, W.E. and C. Shimizu (2015a), "Residential Property Price Indices for Tokyo", *Macroeconomic Dynamics* 19, 1659-1714.

Diewert, W.E. and C. Shimizu (2015b), "A Conceptual Framework for Commercial Property Price Indexes", *Journal of Statistical Science and Application* 3, No.9-10, 131-152.

Diewert, W.E. and C. Shimizu (2017a), "Hedonic Regression Models for Tokyo Condominium Sales," *Regional Science and Urban Economics* 60, 300-315.

Diewert, W.E. and C. Shimizu (2017b), "Alternative Approaches to Commercial Property Price Indexes for Tokyo", *Review of Income and Wealth* 63, 492-519.

Diewert, W.E. and C. Shimizu (2019), "Alternative Land Price Indexes for Commercial Properties in Tokyo", forthcoming in the *Review of Income and Wealth*.

Diewert, W.E. and T.J. Wales (2006), "A 'New' Approach to the Smoothing Problem", pp. 104-144 in *Money, Measurement and Computation*, M.T Belongia and J.M. Binner (eds.), New York: McMillan.

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.

Geltner, D. and S. Bokhari (2015), "Commercial Buildings Capital Consumption in the United States", Final Report, MIT Center for Real Estate, November.

Greville, T.N.E. (1944), "The General Theory of Osculatory Interpolation", *Actuarial Society of America Transactions* 45, 202-265.

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.

Henderson, R. (1924), "A New Method of Graduation", *Actuarial Society of America Transactions* 25, 29-40.

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

Hill, R. J. (2013), "Hedonic Price Indexes for Housing: A Survey, Evaluation and Taxonomy," *Journal of Economic Surveys* 27, 879–914.

Hill, R.J. and M. Scholz (2018), "Can Geospatial Data Improve House Price Indexes? A Hedonic Imputation Approach with Splines", *Review of Income and Wealth 64:4*, 737-756.

Hodrick, R.J. and E.C. Prescott (1980), "Post-War U.S. Business Cycles: An Empirical Investigation", Discussion Paper 451, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Jorgenson, D.W. and Z. Griliches (1967). "The Explanation of Productivity Change", *Review of Economic Studies* 34, 249–283.

Jorgenson, D.W., and Z. Griliches (1972), "Issues of Growth Accounting: A Reply to Edward F. Denison", *Survey of Current Business* 55(5), part II, 65–94.

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.

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.

Mirsky, L. (1955), *An Introduction to Linear Algebra*, Oxford: Clarendon Press.

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

Poirier, D.J. (1976), *The Econometrics of Structural Change*, Amsterdam: North-Holland Publishing Company.

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.

Schoenberg, I.J. (1946), "Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions", *Quarterly Journal of Applied Mathematics* 4, 45-99 and 112-141.

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., 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 (2016), "House Prices at Different Stages of Buying/Selling Process ," *Regional Science and Urban Economics*, 59, 37-53.

Silverman, B.W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting", *Journal of the Royal Statistical Society, Series B*, 47, 1-52.

Sprague, T.B. (1887), "The Graphic Method of Adjusting Mortality Tables", *Journal of the Institute of Actuaries* 26, 77-120.

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.

Wahba, G. (2000), "(Smoothing) Splines in Nonparametric Regression", Technical Report No. 1024, Department of Statistics, University of Wisconsin, 1210 West Dayton St., Madison, WI 53706.

Wahba, G. and J. Wendelberger (1980), "Some New Mathematical Methods for Variational Objective Analysis using Splines and Cross Validation", *Monthly Weather Review* 108, 1122-1143.

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

Whittaker, E.T. (1923), "On a New Method of Graduation", *Proceedings of the Edinburgh Mathematical Society* 41, 63-75.

Wood, S.N. (2004), "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models", *Journal of the American Statistical Association* 99, 673-686.