Econometric Reviews, 34(1–2):106–126, 2015

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2014.944791

Taylor & Francis
Taylor & Francis Group

Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components

W. Erwin Diewert[1], Jan de Haan[2], and Rens Hendriks[3]

[1]*Department of Economics, University of British Columbia, Vancouver, British Columbia, Canada and The School of Economics, University of New South Wales, Sydney, Australia*
[2]*Statistics Netherlands and OTB Research Institute, Delft University of Technology, Delft, Netherlands*
[3]*Statistics Netherlands, The Hague-Leidschenveen, Netherlands*

The paper uses hedonic regression techniques in order to decompose the price of a house into land and structure components using readily available real estate sales data for a small Dutch city. To get sensible results, it was useful to estimate a nonlinear model on data that cover multiple time periods. It also proved necessary to incorporate exogenous information on the rate of growth of construction costs in the Netherlands in order to obtain meaningful constant quality indexes for the price of land and structures separately.

Keywords Fisher ideal indexes; Hedonic regressions; House price indexes; Land and structure components.

JEL Classification C2; C23; C43; D12; E31; R21.

1. INTRODUCTION

For many purposes, it is useful to be able to decompose residential property values into a structures component and a land component. At the local government level, property tax rates are often different on the land and structures components of a property, so it is necessary to have an accurate breakdown of the overall value of the property into these two components. At the national level, statistical agencies need to construct overall values of land and structures for the National Balance Sheets for the nation. If a user cost approach is applied to the valuation of owner-occupied housing services, it is necessary

Address correspondence to W. Erwin Diewert, Department of Economics, University of British Columbia, Vancouver, B.C. V6T 1Z1, Canada; E-mail: diewert@econ.ubc.ca

to have a decomposition of housing values into land and structures components since structures depreciate while land does not.

The purpose of this paper is to decompose the selling price of each property into a land component and a structures component using readily available multiple listing data on sales of residential properties. Basically, variations of *three methods* can be found in the literature to provide such a decomposition: the vacant land method, the construction cost method, and hedonic regression methods. The first two methods utilize the following empirical relationship between the selling price of a property V, the value of the structure $p_S S$, and the value of the plot $p_L L$:

$$V = p_L L + p_S S, \tag{1}$$

where S is the floor space area of the structure, L is the area of the land that the structure sits on, and p_S and p_L are the prices of a unit of S and L, respectively. Typically, V, L, and S will be available from real estate data on sales of houses, so if either p_L or p_S can be determined somehow, then Eq. (1) will enable the other price to be determined.

The *vacant land method* for the determination of the price of land in (1) is described by Clapp (1979, 1980), who noted that it is frequently used by tax assessors and appraisers. The method works as follows. A price of land per unit area p_L is determined from the sales of "comparable" vacant land plots, and this price is applied to comparable properties. Equation (1) can then be used to solve for the structure price p_S. This method was used by Thorsnes (1997) and Bostic et al. (2007). The set of vacant lots can be augmented by properties which are sold and the associated structure is immediately demolished. Clapp (1980) lists a number of reasons why the vacant land method is not likely to be very accurate.

The *construction cost method* uses an estimate for the per unit area construction cost p_S for the local area, which could be provided by a private company or a national statistical agency. Once p_S is known, Eq. (1) can be used to solve for the missing land price p_L. This method was used by Glaeser and Gyourko (2003), Gyourko and Saiz (2004), and Davis and Palumbo (2008); the local construction cost data for U.S. cities was provided by the private company, R.S. Means. Davis and Heathcote (2007) used a variant of this method for the entire U.S. economy, where Bureau of Economic Analysis estimates for both the price of structures p_S and the constant dollar quantity of housing structures S were used.[1]

[1] Muth (1971) and Rosen (1978) used the private company Boeckh building cost index for the various U.S. cities in their sample which determined p_S up to a multiplicative factor. The value of land and the price of land were determined by the U.S. Federal Housing Administration for their sample of U.S. properties. Then using Eq. (1), S was determined residually. The methods we will use in Sections 4 and 5 below are close to the construction cost method but are not identical; we use only rates of change of construction costs, not their levels. Thus our suggested methods allow for local area quality adjustment factors for construction costs.

A variant of the *hedonic regression method* will be used in this paper. Some early papers that use a similar methodology include Clapp (1980), Palmquist (1984), Fleming and Nellis (1992), and Schwann (1998). Land and structures are treated as characteristics in a hedonic regression model, and marginal prices for land and structures for period t are generated as partial derivatives of the period t hedonic function. These marginal prices can be used to decompose the house value into land and structures components under certain conditions.

Our approach is based on a cost oriented model which we call the *builder's approach* to modeling hedonic regressions in the housing context. A distinct feature is that it requires relatively little information on the characteristics of the houses. Using data for 22 quarters pertaining to detached houses in a small Dutch city, we show that information on the plot area, the area of the structure, the age of the structure, and the number of rooms suffices to generate regression models that explain approximately 87% of the variation in the selling prices. A more detailed outline of the contents of this paper follows.

In Section 2, we consider a very simple hedonic model that has only three characteristics of the property as explanatory variables: lot size, size of the structure, and (approximate) age of the structure. We postulate that the value of a residential property is the *sum* of the value of the land which the structure sits on and the value of the residential structure. Thus, our approach to the valuation of a residential property is essentially a crude cost of production approach. We run a separate hedonic regression for each quarter which leads to quarterly estimates of the prices for land and structures. These estimated characteristics prices can then be converted into land and structures price indexes for the 22 quarters of data in our sample.

In Section 3, we generalize the model explained in Section 2 to allow for the observed fact that the per unit area price of a property tends to decline as the size of the lot increases (at least for large lots). We use a simple linear spline model with two break points. Again, a hedonic regression is run for each period and the results of these separate regressions were linked together to provide separate land and structures price indexes along with an overall price index that combined these two components.

The models described in Sections 2 and 3 were not successful, due to *multicollinearity* and *variability* in the data. This data volatility leads to a tendency for the regression models to fit the outliers, leading to erratic estimates for the price of land and structures. To deal with the multicollinearity problem, in Section 4 we draw on *exogenous information* on new house building costs from the national statistical agency and assume that the price *movements* for new structures in our data set mirror the officially measured nationwide movements. We find that the use of exogenous information produces a very reasonable decomposition of house values into their structure and land components.

In Section 5, we generalize the model in Section 4 to include information on the number of rooms in the house as an additional price determining characteristic. The idea is that a larger number of rooms generally indicates that the quality of construction of the house will be higher. Our regression results support this hypothesis.

2. A SIMPLE BUILDER'S MODEL

Hedonic regression models are frequently used to obtain constant quality price indexes for owner occupied housing.[2] A residential property has a number of important price determining characteristics, such as the land area of the property; the livable floor space area of the structure; the age of the structure; the number of rooms; the type of dwelling unit (detached, row, apartment), the type of construction (wood, brick, concrete), and the location of the property.[3] In our empirical work below, we will restrict our sample to sales of detached houses. We will not take into account the type of construction or the location. The sales all take place in a small Dutch town, and location should not be much of a price-determining factor. However, we will use information on land area, structure size in meters squared, and age of the structure.

Several researchers have suggested the use of hedonics to decompose the overall price of a property into *additive components* that reflect the value of the land the structure sits on and the value of the structure.[4] We outline Diewert's (2007) justification for an additive decomposition. Consider a property developer who is planning to build 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 equal 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 V_n^t in period t[5] and structure areas S_n^t and land areas L_n^t for $n = 1, \ldots, N(t)$. Assume that these prices are equal to the sum of the land and structure costs plus error terms ε_n^t which we assume are independently normally distributed with zero means and constant variances.[6] This leads to the following hedonic regression model for period t, where α^t and β^t are the

[2]For some recent literature, see Crone et al. (2009), Diewert et al. (2009), Gouriéroux and Laferrère (2009), Hill (2011), and Hill et al. (2009). One variant of the hedonic technique regresses the logarithm of the selling price of the property on its price determining characteristics and on time dummy variables for all periods (except the base period). The estimated time dummy coefficients can be exponentiated and turned into an index. In another approach, separate hedonic regressions are estimated for each of the periods compared; this is called the hedonic imputation approach. See Haan (2009, 2010) and Diewert et al. (2009) for theoretical discussions and comparisons between these alternative approaches.

[3]There are many other price determining characteristics that could be added to this list such as landscaping, the number of floors, type of heating system, air conditioning, swimming pools, views, the shape of the lot, etc. The distance of the property to various amenities such as schools and shops could also be added to the list of characteristics.

[4]See Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic et al. (2007), Davis and Heathcote (2007), Diewert (2007), Francke (2008), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert (2010), and Diewert et al. (2011).

[5]We have labeled these property prices as V_n^t to emphasize that these are *values* of the property, and we need to decompose these values into two price and two quantity components, where the components are land and structures.

[6]We make the same stochastic assumptions for all of the regressions in this paper. For the models that are not linear in the parameters, we use maximum likelihood estimation.

parameters to be estimated in the regression:[7]

$$V_n^t = \alpha^t L_n^t + \beta^t S_n^t + \varepsilon_n^t; \qquad n = 1, \ldots, N(t); \qquad t = 1, \ldots, T. \tag{2}$$

Note that the two characteristics in our simple model are the quantities of land L_n^t and the quantities of structure S_n^t associated with the sale of 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 (2) for each period t in our sample.

The hedonic regression model defined by (2) is the simplest possible one, but it applies only to new structures. However, it is likely that a model that is similar to (2) applies to older structures as well. Older structures will usually be worth less than newer structures due to depreciation (or deterioration due to aging effects) of the structure. Suppose in addition to information on the selling price of property n at time period t, V_n^t, the land area of the property L_n^t and the structure area S_n^t, we also have information on the age of the structure at time t, say A_n^t. Then, if we assume a straight line depreciation model, a more realistic hedonic regression model than that defined by (2) above is the following *basic builder's model*:[8]

$$V_n^t = \alpha^t L_n^t + \beta^t (1 - \delta^t A_n^t) S_n^t + \varepsilon_n^t; \qquad n = 1, \ldots, N(t); \qquad t = 1, \ldots, T, \tag{3}$$

where the reflects the net depreciation rate as the structure ages one additional period. If the age of the structure is measured in years, we would expect an annual δ^t to be between 0.5 and 1.5%. This parameter δ^t is regarded as a *net depreciation rate* because it is equal to a "true" gross structure depreciation rate less an average renovations appreciation rate. Since we do not have information on renovations and additions to a structure at our disposal, our age variable will only pick up average gross depreciation less average real renovation expenditures. 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 their inclusion can bias the estimates of the net depreciation rate for younger structures.

[7]In order to obtain homoskedastic errors, it would be preferable to assume multiplicative errors in equation (2), since it is more likely that expensive properties have relatively large absolute errors compared to very inexpensive properties. However, following Koev and Santos Silva (2008), we think that it is preferable to work with the additive specification (2) because we are attempting to decompose the aggregate value of housing into additive structures and land components, and the additive error specification will facilitate this decomposition.

[8]This model is a *supply side model* as opposed to the *demand side model* of Muth (1971) and McMillen (2003). Basically, we are assuming identical suppliers of housing so that we are in Rosen's (1974) Case (a) where the hedonic surface identifies the structure of supply. This assumption is justified for the case of newly built houses, but we concede that it is less well justified for sales of existing homes. Our supply side model is also less likely to be applicable in the case of multiple unit structures where zoning restrictions and local geography lead to location specific land prices.

Note that (3) is a nonlinear regression model whereas (2) was a simple linear regression model.[9] Both Eqs. (2) and (3) can be estimated period by period; it is not necessary to run one big regression covering all time periods in the data sample. The period t price of land will the estimated coefficient for the parameter α^t and the price of a unit of a newly built structure for period t will be the estimate for β^t. The period t quantity of land for property n is L_n^t and the period t quantity of structure for property n, expressed in equivalent units of a new structure, is $(1 - \delta^t A_n^t)S_n^t$, where S_n^t is the floor space area of property n in period t.

We estimated the above *Model 0* and subsequent models using real estate data on the sales of detached houses for a small city (population is around 60,000) in the Netherlands, City "A," for 22 quarters, starting in the first quarter 1 of 2003 and extending through the second quarter of 2008 (so our $T = 22$). The data exploited can be described as follows:

1. V_n^t is the selling price of property n in quarter t in units of 1,000 Euros where $t = 1, \ldots, 22$;
2. L_n^t is the area of the plot for the sale of property n in quarter t in units of meters squared;[10]
3. S_n^t is the living space area of the structure for the sale of property n in quarter t in units of meters squared;
4. A_n^t is the (approximate) age in decades of the structure on property n in period t;[11]
5. R_n^t is the number of rooms in structure n that was sold in period t.

It seems likely that the number of rooms in a structure will be roughly proportional to the area of the structure, so in our initial regressions, we did not use the room variable as an explanatory variable.[12]

Initially, there were 3,543 observations in our 22 quarters of data on sales of detached houses in City "A" that were less than 50 years old when sold. However, there were some obvious outliers in the data. We looked at the range of our V, L, S, and R variables and deleted 54 range outliers. There were two duplicate data points in Q1 for 2006, and these duplicates were also deleted. We ended up with 3,487 observations for the 22 quarters.[13]

[9]This formulation follows that of Diewert (2007) and Diewert et al. (2011). It is a special case of Clapp's (1980) model except that Clapp included a constant term.

[10]We chose units of measurement for V in order to scale the data to be small in magnitude so as to facilitate convergence for the nonlinear regressions. The statistical package used for estimating all models in this paper was Shazam (the nonlinear option).

[11]The original data were coded as follows: if the structure was built 1960–1970, the observation was assigned the dummy variable BP = 5; 1971–1980, BP = 6; 1981–1990, BP = 7; 1991–2000, BP = 8. Our Age variable A was set equal to 8 − BP. Thus for a recently built structure n in quarter t, $A_n^t = 0$.

[12]In Section 5 below, we will use the number of rooms as an additional quality adjustment variable.

[13]There were 3 observations where the selling price was less than 60,000 and 14 observations which sold for more than 550,000 Euros. There were no sales with L less than 70 m^2 and 25 sales where L exceeded 1,500 m^2. There were no sales with S less than 50 and one observation where S exceeded 400 m^2. There were 13 sales where R was less than 2 and 3 sales where R exceeded 14. All of these observations were excluded. Some observations were excluded multiple times so that the total number of observations which were excluded was 54 (plus 2 more due to duplication in the data set). Exclusion of range outliers is important for the results.

The sample means for the data with outliers excluded (standard deviations in brackets) were $\overline{V} = 182.26(71.3)$, $\overline{L} = 258.06(152.3)$, $\overline{S} = 126.56(29.8)$, $\overline{A} = 1.8945(1.23)$, and $\overline{R} = 4.730(0.874)$. The sample median price was 160,000 Euros.

The correlations between the variables are also of interest. The correlation coefficients of the selling price V with L, S, A, and R are 0.8014, 0.7919, −0.3752, and 0.3790, respectively.[14] Thus, the selling price V is fairly highly correlated with both land L and (unadjusted) structures S. The correlation between L and S is 0.6248, which points to the possibility of a multicollinearity problem when estimating the regressions. Finally, there is a substantial positive correlation of 0.4746 between the structure area S and the number of rooms R.

Instead of running 22 quarterly regressions of the form (3), we combined the data using dummy variables and ran one big regression, which combined all 22 quarterly regressions into a single regression.[15] The R^2 for the resulting combined regression was 0.8729, which is quite good, considering we have only 3 explanatory variables (but 66 parameters to estimate).[16] The log likelihood was $-16,231.6$. The *quality adjusted structures quantity in quarter t*, S^{t^*}, is equal to the sum over the properties sold in that quarter adjusted into new structure units; i.e., $S^{t^*} \equiv \sum_{n\in N(t)} (1 - \delta^{t^*} A_n^t)S_n^t$. The estimated decade net depreciation rates δ^{t^*} were in the 6.4% to 13.7% range, which is not unreasonable, but the volatility in these rates is not consistent with our a priori expectation of a stable depreciation rate.

In an attempt to improve the results for the above Model 0, we assumed that the net depreciation rate was constant across quarters and so the model defined by (3) is replaced by the following *Model 1*:

$$V_n^t = \alpha^t L_n^t + \beta^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t; \qquad n = 1, \ldots, N(t); \qquad t = 1, \ldots, T, \tag{4}$$

where the parameter δ reflects the sample *net depreciation rate* as the structure ages one additional decade, which is now assumed to be constant over the entire sample period. The new builder's hedonic regression model has 45 unknown parameters to estimate as compared to the 66 parameters in the previous model defined by Eq. (3).

The R^2 for the resulting nonlinear regression model was 0.8703, which is quite good, considering we have only two independent explanatory variables in each period. However, this is a drop in R^2 as compared to our previous model with variable depreciation rates where the R^2 was 0.8729. The log likelihood for the constant depreciation rate model was −16,266.6, which is a decrease of 35.0 from the log likelihood of the previous model. This

[14] The correlation coefficients of V with L, S, A, and R for the original data set with 3,543 observations were 0.33331, 0.80795, −0.34111, and 0.34291. This illustrates the importance of deleting land outliers in particular.

[15] This big regression generates the same parameter values as running the individual quarterly regressions, but the advantage of the one big regression approach is that we can compare the log likelihood of the big regression with subsequent regressions.

[16] All the R^2 values 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.

decrease in log likelihood seems to be a reasonable price to pay in order to obtain a stable estimate for the net depreciation rate. The estimated decade net depreciation rate is now $\delta^* = 0.10241$ or about 1% per year. Using this "fixed" value for the net depreciation rate, the *quality adjusted structures quantity in quarter t*, S^{t^*}, is now estimated as

$$S^{t^*} \equiv \sum_{n \in N(t)} (1 - \delta^* A_n^t) S_n^t; \qquad t = 1, \ldots, 22. \tag{5}$$

The estimated prices of land and structures, α^{t^*} and β^{t^*}, were normalized to equal 1 in quarter 1 and are named P_{L1} and P_{S1}. We used α^{t^*} and β^{t^*}, along with the corresponding quantity data, L^t and S^{t^*}, to calculate a constant quality chained Fisher (1922) house price index, labeled P_1. The land and structures price series and the overall Fisher index are graphed in Fig. 1. The changes in the sample mean and median prices (normalized to equal 1 in quarter 1), P_{Mean} and P_{Median}, are also shown.

The mean and median price series are on average substantially above the Fisher house price index P_1; P_1 is also much smoother. We attribute the slower rate of growth in the hedonic index to the fact that new houses tend to get bigger over time; the mean and median indexes do not adjust for this quality improvement. However, while P_1 appears to provide satisfactory results for the overall house price change, the component land and structure price series, P_{L1} and P_{S1}, are extremely volatile: when the price of land spikes up, the price of structures tends to spike down, and vice versa. This erratic behavior in P_{L1} and P_{S1} is a sign of severe multicollinearity due to the high correlation between the

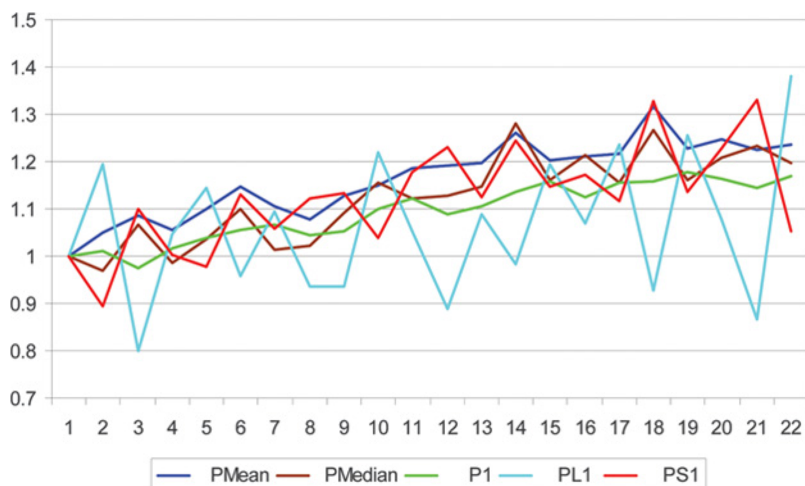

FIGURE 1 Mean and median price indexes, land price index P_{L1}, structures price index P_{S1}, and overall house price index P_1.

quantity of land and structures.[17] Measurement errors in the quantity of land and the quantity of structures may also explain part of the volatility. These include both recording errors and errors due to our imperfect measurement of the quality of construction and the quality of the land. For example, we are assuming that all locations in our sample have access to the same amenities and share the same geography and hence should face the same land price schedule, but this will not be completely true.

The multicollinearity problem will be addressed in Section 4. One other problem with our highly simplified house price model is that it makes no allowance for the fact that larger sized plots typically sell for a lower average price than medium and smaller sized plots.[18] In the following section, we will generalize our builder's model to take into account this empirical regularity.

3. THE BUILDER'S MODEL WITH LINEAR SPLINES ON LOT SIZE

In most countries, including the Netherlands, large lots tend to sell at a lower price per unit area than smaller lots. We assume that builders face a piecewise linear schedule of prices per unit land when they purchase a lot. This *linear spline model* allows the price of large lots to drop as compared to smaller lots. We broke up our 3,487 observations into three groups of roughly equal size: sales involving lot sizes less than 170 meters squared (group S), sales involving lot sizes between 170 and less than 270 meters squared (group M), and sales involving lot sizes greater than or equal to 270 meters squared (group L). We denote the sets of observations n which belong to group S, M, and L in period t by $N_S(t)$, $N_M(t)$, and $N_L(t)$, respectively.

For an observation n in period t that was associated with a small lot size, our regression model was essentially the same as (4); i.e., the following estimating equation was used:

$$V_n^t = \alpha_S^t L_n^t + \beta^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t; \qquad t = 1, \ldots, 22; \qquad n \in N_S(t), \tag{6}$$

where the unknown parameters to be estimated are α_S^t, β^t for $t = 1, \ldots, 22$ and δ. For an observation associated with a medium lot size, the estimating equation was

$$V_n^t = \alpha_S^t (170) + \alpha_M^t (L_n^t - 170) + \beta^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t; \qquad t = 1, \ldots, 22; \qquad n \in N_M(t), \tag{7}$$

[17]A similar parameter instability problem was noted by Schwann (1998, p. 277) in his initial unconstrained model: "In addition, the unconstrained regression displays signs of multicollinearity.... the attribute prices are nonsense in many of the periods, and there is poor temporal stability of these prices."

[18]This empirical regularity was noted by Francke (2008, p. 168): "However, the assumption that the value is proportional to the lot size is not valid for large lot sizes. In practice, real estate agents often use a step function for the valuation of the lot...." At first glance, it seems that Francke used a step function to model the price schedule, but it turns out that he used linear splines in the same way as we do.

where we added 22 new parameters to be estimated, the α_M^t for $t = 1, \ldots, 22$. Finally, for an observation associated with a large lot size, the following estimating equation was used:

$$V_n^t = \alpha_S^t(170) + \alpha_M^t(270 - 170) + \alpha_L^t(L_n^t - 270) + \beta^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t;$$
$$t = 1, \ldots, T; \qquad n \in N_L(t), \tag{8}$$

where we have added 22 new parameters to be estimated, the α_L^t for $t = 1, \ldots, 22$. Thus for small lots, the value of an extra marginal addition of land in quarter t is α_S^t, for medium lots, the value of an extra marginal addition of land in quarter t is α_M^t, and for large lots, the value of an extra marginal addition of land in quarter t is α_L^t. These pricing schedules are joined together so that the cost of an extra unit of land increases with the size of the lot in a continuous fashion.[19] *Model 2* defined by equations (6)–(8) was estimated on the total sample using dummy variables to indicate whether an observation is in group S, M, or L.

The R^2 for this model was 0.8756, a small increase over the previous two models (without splines), where the R^2 was 0.8729 (many depreciation rates) and 0.8703 (one depreciation rate). The log likelihood was −16,195.0, an increase of 71.6 from the previous model's log likelihood. The estimated decade depreciation rate was $\delta^* = 0.1019(0.00329)$.[20] The first period parameter values for the three marginal prices for land were $\alpha_S^{1*} = 0.2889(0.0497)$, $\alpha_M^{1*} = 0.3643(0.0566)$, and $\alpha_L^{1*} = 0.1895(0.319)$. Thus in quarter 1, the marginal cost per m^2 was estimated to be 288.9 Euros, 364.3 Euros, and 189.5 Euros, respectively, for small lots, medium sized lots, and large lots. The first period parameter value for quality adjusted structures was $\beta^{1*} = 0.8829(0.0800)$, so that a square meter of new structure was valued at 882.9 Euros. All of the estimated coefficients were positive, as expected, and were significantly different from zero. Our conclusion is that adding linear splines for the lot size gives us some additional explanatory power.

Next, we calculated the predicted value of land for small, medium, and large lot sales in each quarter t, V_{LS}^t, V_{LM}^t, and V_{LL}^t, and the associated quantities of land, L_{LS}^t, L_{LM}^t, and L_{LL}^t, as follows:

$$V_{LS}^t \equiv \sum_{n \in N_S(t)} \alpha_S^{t*} L_n^t; \qquad t = 1, \ldots, 22; \tag{9}$$

$$V_{LM}^t \equiv \sum_{n \in N_M(t)} \alpha_S^{t*}[170] + \alpha_M^{t*}[L_n^t - 170]; \qquad t = 1, \ldots, 22; \tag{10}$$

[19]If we graphed the total cost C of a lot as a function of the plot size L in period t, the resulting cost curve would be made up of three linear segments whose endpoints are joined. The first line segment starts at the origin and has the slope α_S^t, the second segment starts at $L = 170$ and runs to $L = 270$ and has the slope α_M^t, and the final segment starts at $L = 270$ and has the slope α_L^t.

[20]Standard errors are in brackets.

$$V_{LL}^t \equiv \sum_{n \in N_L(t)} \alpha_S^{t^*}[170] + \alpha_M^{t^*}[100] + \alpha_L^{t^*}[L_n^t - 270]; \qquad t = 1, \ldots, 22; \tag{11}$$

$$L_{LS}^t \equiv \sum_{n \in N_S(t)} L_n^t; \qquad t = 1, \ldots, 22; \tag{12}$$

$$L_{LM}^t \equiv \sum_{n \in N_M(t)} L_n^t; \qquad t = 1, \ldots, 22; \tag{13}$$

$$L_{LL}^t \equiv \sum_{n \in N_L(t)} L_n^t; \qquad t = 1, \ldots, 22. \tag{14}$$

The corresponding *average prices*, P_{LS}^t, P_{LM}^t, and P_{LL}^t, are defined as the above values divided by the above quantities:

$$P_{LS}^t \equiv V_{LS}^t / L_{LS}^t; \qquad P_{LM}^t \equiv V_{LM}^t / L_{LM}^t; \qquad P_{LL}^t \equiv V_{LL}^t / L_{LL}^t; \qquad t = 1, \ldots, 22. \tag{15}$$

The average land prices for the tree types of lot defined by (15) and the corresponding quantities of land defined by (12)–(14) were to form a chained *Fisher land price index*, which we denote by P_{L2}. As in the previous model, the estimated price for a square meter of quality adjusted structures in quarter t is β^{t^*}, and the corresponding quantity of constant quality structures is $S^{t^*} \equiv \sum_{n \in N(t)} (1 - \delta^* A_n^t) S_n^t$. The structure price and quantity series β^{t^*} and S^{t^*} were combined with the three land price and quantity series to form a chained *overall Fisher house price index* P_2. The *constant quality structures price index* P_{S2} is again a normalization of the series $\beta^{1^*}, \ldots, \beta^{22^*}$.

Figure 2 compares the new price series P_{L2}, P_{S2}, and P_2 with the price series P_{L1}, P_{S1}, and P_1 generated by Model 1 (without splines on the size of the land area). It can be seen that Models 1 and 2 generate essentially the same overall house price series; P_1 and P_2 can hardly be distinguished. The two overall series are quite smooth and look reasonable. But the multicollinearity problem with the price of land and the price of structures remains. In fact, the offsetting jumps using the splines Model 2 are bigger than they were using the no splines Model 1.

One possible approach to eliminating the volatility problem is to use a *smoothing method* in order to stabilize the period to period characteristics prices.[21] We have not pursued this approach because we feel that it is not an appropriate one for statistical agencies who have to produce nonrevisable housing price indexes in real time. The use of smoothing methods is appropriate when the task is to produce historical series, but these techniques do not work well in a real time context due to their inability to predict turning points in the series.

[21]This approach dates back to Coulson (1992) and Schwann (1998). More recent contributions include Francke and Vos (2004), Francke (2010), and Rambaldi et al. (2011).

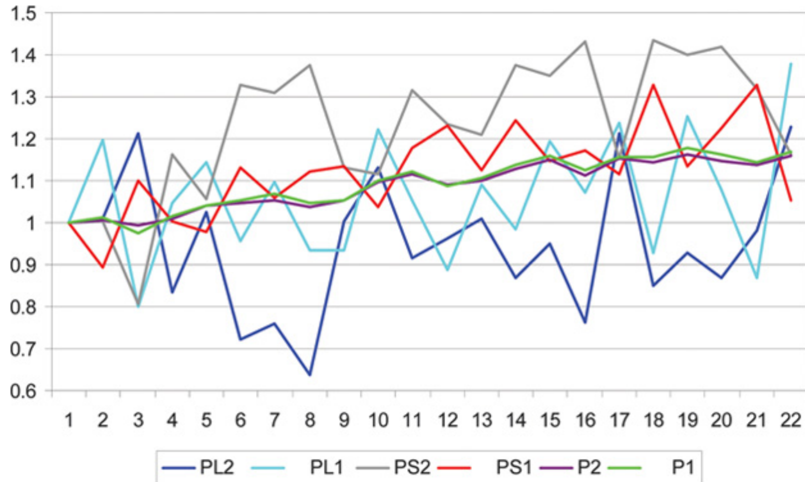

FIGURE 2 Land price indexes P_{L1} and P_{L2}, structures price indexes P_{S1} and P_{S2}, and overall house price indexes P_1 and P_2.

What we did instead to resolve the multicollinearity problem and obtain meaningful land and structures price series was adding some *additional restrictions* on Model 2 by using exogenous information on new construction prices.[22] This is the topic discussed in Section 4.

4. THE USE OF EXOGENOUS INFORMATION ON NEW CONSTRUCTION PRICES

Many countries, including the Netherlands, have national (or regional) new construction price indexes available from the national statistical agency on a quarterly basis.[23] We make the assumption that new construction costs for houses have the same rate of growth over the sample period across all cities in the Netherlands and show how the statistical agency information on construction costs can be used to eliminate the multicollinearity problems encountered in the previous sections.[24]

[22]In an earlier paper (Diewert et al., 2011), we assumed that the price of structures never decreased, and using this assumption led to a reasonable decomposition of overall price change into land and structure components. However, for our new sample period, it appears that this assumption is not warranted and so a different approach will be used in the present paper.

[23]As was seen in Section 1, many countries have private companies that can provide timely construction price indexes for major cities in the country, and this information could also be used.

[24]From Statistics Netherlands' online source, Statline, we obtained a quarterly series for "New Dwelling Output Price Indices, Building Costs, 2005 = 100, Price Index: Building costs including VAT" for the last 14 quarters in our sample. Data from Statline for the first 8 quarters in our sample were also available but using the base year 2000 = 100. The older series was linked to the newer series, and the resulting series was normalized to 1 in the first quarter. The resulting series is denoted by $p^1(= 1), p^2, \ldots, p^{22}$.

Recall the estimating Eqs. (6)–(8) for Model 2 in Section 3. We replace the constant quality structures parameters β^t for $t = 2, \ldots, 22$ by the following numbers, which involve only the single unknown parameter β^1:

$$\beta^t = \beta^1 p^t; \qquad t = 2, 3, \ldots, 22, \tag{16}$$

where p^t is the statistical agency estimated *construction cost price index* for the location under consideration and for the type of dwelling, where this series has been normalized to equal 1 in quarter 1. This new regression *Model 3* is again defined by Eqs. (6)–(8) except that the 22 unknown β^t parameters are now assumed to be defined by (16), so that only β^1 needs to be estimated for this new model.[25] The total number of parameters to be estimated in this restricted model is 68, while in Model 2 it was 89.

Using the data for the City "A," the estimated decade depreciation rate was $\delta^* = 0.1026(0.00448)$. The R^2 for this model was 0.8723, a drop compared to the value of 0.8756 for Model 2. The log likelihood was −16,239.7, a substantial decrease of 44.7 over Model 2. The first quarter estimates for the three marginal prices for land were $\alpha_S^{1*} = 0.1827(0.0256)$, $\alpha_M^{1*} = 0.3480(0.0640)$, and $\alpha_L^{1*} = 0.17064(0.0311)$. The first quarter estimate for the price of quality adjusted structures was $\beta^{1*} = 1.0735(0.0275)$ or 1,073.5 Euros/m^2, which exceeds Model 1 and 2 estimates of 972.1 and 882.9 Euros/m^2, respectively. Thus, the imposition of a nationwide growth rate on the change in the price of quality adjusted structures for the city of "A" has had some effect on our previous estimates for the *levels* of land and structures prices.

We used Eqs. (9)–(15) to construct a chained Fisher index of land prices, which we label P_{L3}. As for the previous two models, the estimated period t price for a square meter of quality adjusted structures is denoted by β^{t*} (which in turn is now equal to β^{1*} p^t), and the corresponding quantity of constant quality structures is $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \delta^* A_n^t) S_n^t$. The structures price and quantity series β^{t*} and S^{t*} were again combined with the three land price and quantity series to form a chained overall Fisher house price index P_3. The constant quality structures price index P_{S3} (a normalization of the series $\beta^{1*}, \ldots, \beta^{22*}$). It should be noted that the quarter to quarter movements in P_{S3} coincide with the quarter to quarter movements in Statistics Netherlands' New Dwellings Building Cost Price Index. Table 1 contains the various price indexes and the corresponding quantities generated by Model 3.

Figure 3 plots the price series listed in Table 2, P_{L3}, P_{S3}, and P_3, along with the overall house price indexes generated by Models 1 and 2, P_1, and P_2, It can be seen that P_1, P_2, and P_3 can barely be distinguished as separate series in Fig. 3.[26] The price of structures

[25]This type of hedonic model is similar to that introduced by Diewert (2010). Note that our present model is close to the construction cost method explained in Section 1, but it is not identical since we still estimate a large number of spline parameters plus the structures quality parameter β^1 and the net depreciation rate δ.

[26]We ran a wide variety of hedonic regressions using the same price and characteristics data but different functional forms for the various regressions and found that they all fitted the overall price data fairly well and generated similar *overall* house price indexes. However, these alternative models did not generate reasonable land and structures price indexes.

TABLE 1
Land Price Index P_{L3}, Structures Price Index P_{S3}, Overall House Price Index P_3, and the Corresponding Quantities Q_{L3}, Q_{S3}, and Q_3

Quarter	P_{L3}	P_{S3}	P_3	Q_{L3}	Q_{S3}	Q_3
1	1.0000	1.0000	1.0000	7,447	15,749	23,196
2	0.9925	1.0161	1.0084	7,602	15,074	22,671
3	0.9925	1.0000	0.9977	8,623	15,752	24,366
4	1.0440	0.9919	1.0104	9,173	17,988	27,139
5	1.1479	0.9839	1.0401	8,058	15,868	23,904
6	1.2096	0.9597	1.0455	9,899	18,058	28,027
7	1.2244	0.9677	1.0559	7,200	14,201	21,364
8	1.1116	1.0000	1.0406	8,659	18,424	26,956
9	1.2013	0.9839	1.0582	8,286	17,810	26,049
10	1.3590	0.9769	1.1043	8,221	17,006	25,162
11	1.3649	0.9988	1.1210	8,406	16,988	25,373
12	1.2492	1.0227	1.0981	8,843	17,298	26,170
13	1.3316	0.9908	1.1050	9,339	18,107	27,488
14	1.4058	1.0008	1.1365	9,931	20,429	30,275
15	1.4719	0.9958	1.1549	8,437	17,051	25,455
16	1.3527	0.9988	1.1171	8,633	16,994	25,649
17	1.4476	1.0177	1.1614	8,567	17,421	25,945
18	1.3948	1.0277	1.1498	12,263	22,083	34,613
19	1.4018	1.0516	1.1677	7,709	14,659	22,456
20	1.3205	1.0745	1.1555	10,337	21,044	31,382
21	1.2561	1.0954	1.1483	8,142	16,466	24,615
22	1.3114	1.0954	1.1663	9,854	21,082	30,881
Mean	1.2541	1.0114	1.0928	8,801	17,529	26,324

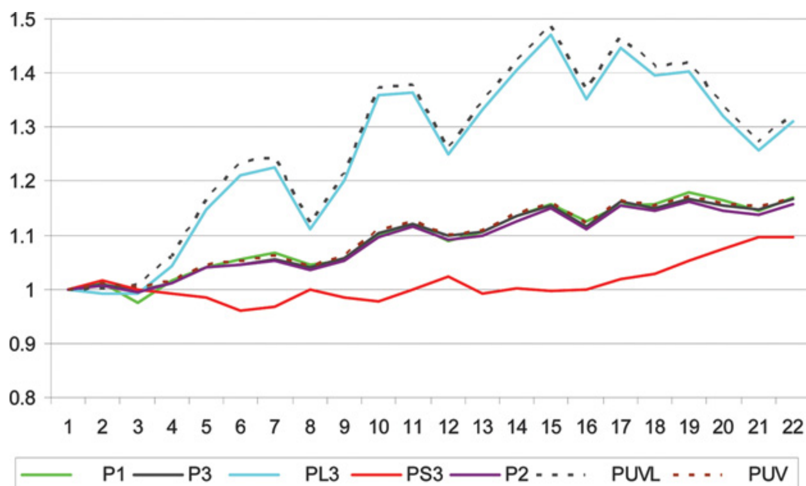

FIGURE 3 Land price index P_{L3}, structures price index P_{S3}, overall house price indexes P_1, P_2, and P_3 and a unit value land price index P_{UVL} and overall index P_{UV}.

TABLE 2
Land Price Index P_{L4}, Structures Price Index P_{S4}, Overall House Price Index P_4, and the Corresponding Quantities Q_{L4}, Q_{S4}, and Q_4

Quarter	P_{L4}	P_{S4}	P_4	Q_{L4}	Q_{S4}	Q_4
1	1.0000	1.0000	1.0000	8,373	14,816	23,189
2	0.9892	1.0161	1.0063	8,499	14,218	22,713
3	0.9825	1.0000	0.9936	9,541	14,929	24,459
4	1.0318	0.9919	1.0076	10,216	17,006	27,202
5	1.1289	0.9839	1.0390	8,980	14,954	23,918
6	1.1848	0.9597	1.0456	10,954	17,004	28,021
7	1.1979	0.9677	1.0556	8,002	13,398	21,364
8	1.1015	1.0000	1.0407	9,691	17,364	26,943
9	1.1745	0.9839	1.0563	9,264	16,952	26,090
10	1.3187	0.9769	1.1037	9,172	16,053	25,167
11	1.3233	0.9988	1.1193	9,386	16,036	25,406
12	1.2195	1.0227	1.0956	9,832	16,368	26,222
13	1.3026	0.9908	1.1072	10,381	17,003	27,430
14	1.3615	1.0008	1.1353	11,027	19,376	30,305
15	1.4193	0.9958	1.1533	9,407	16,110	25,486
16	1.3085	0.9988	1.1141	9,592	16,114	25,713
17	1.3905	1.0177	1.1563	9,544	16,562	26,054
18	1.3381	1.0277	1.1427	13,606	21,006	34,826
19	1.3537	1.0516	1.1633	8,590	13,876	22,540
20	1.2863	1.0745	1.1524	11,516	19,960	31,465
21	1.2223	1.0954	1.1422	9,076	15,670	24,740
22	1.2828	1.0954	1.1641	11,010	19,981	30,934
Mean	1.2236	1.0114	1.0906	9,803	16,580	26,372

does not behave in a monotonic manner; after dipping 5% in quarter 6, it trends up to finish about 10% higher at the end of the sample period as compared to the beginning of the sample period. The price of land P_{L3} peaked in quarter 15, approximately 47% higher than the quarter 1 level, and then it trended downwards to finish 31% higher in quarter 22. According to Model 3, the price of land fluctuates much more than the price of structures. This is exactly what we would expect to find, so this model seems to perform reasonably well.

It is useful to compare our Model 3 with an *approximate construction cost model of land and structure prices*. In order to do this, we require a net depreciation rate and period-by-period levels of construction costs. We will use our estimated Model 3 net depreciation rate of $\delta^* = 0.1026$ and our estimated starting level of construction costs per square meter of 1,073.5 Euros/m^2, and then we use the Statistics Netherlands index of construction costs to construct period-by-period construction cost estimates per meter squared. Using these estimates, we calculated the aggregate value of structures sold during each period, and then we subtracted this estimated structures value from the total value of housing sales, leading to an estimate for the value of land sold during the period. A unit value price for land for each period was then constructed by dividing the estimated land value by the total land area of the sales for that period. The resulting unit value land price P_{UVL} is graphed in Fig. 3 and is compared with our Model 3 hedonic regression model price of land P_{L3}. It can be seen that these two indexes capture the same trends. We also constructed a Fisher chained aggregate of P_{UVL} and P_{S3} and called the resulting overall index P_{UV}. It is compared with our preferred overall index P_3 in Fig. 3, and it can be seen that these two indexes are very close. Our conclusion at this point is that the construction cost method for constructing separate land and structures price indexes can work rather well in capturing overall price trends. However, we think that our hedonic approach can potentially construct more accurate price indexes by bringing additional housing characteristics into the model, thus leading to better quality adjustments and less unit value bias. In the following section, we show how extra housing characteristics can be introduced into our model.

5. THE USE OF ADDITIONAL CHARACTERISTICS INFORMATION

In the last two models, we made use of the fact that large lots are likely to have a lower price per meter squared than medium lots. By modeling this empirical regularity with the use of splines on the quantity of land, we were able to improve the fit of the regression. It is also likely that larger structures have a higher quality than small structures due to the use of more expensive construction materials. Thus, it seems quite natural to use the same type of spline setup, but on structures S rather than land L, in order to further improve our model.

However, a more parsimonious alternative to using spline techniques on structures is to utilize information on the number of rooms in the structure; i.e., as the number of rooms increases, we would expect the quality of the structure to increase so that the price per meter squared of a structure is expected to increase as the number of rooms increases.[27] Below, we will incorporate the room variable R in our hedonic model. It should be noted, though, that some housing experts believe that the price per meter squared should decline as the structure size increases, so the issue is not settled.[28]

Our new regression *Model 4* is defined by Eqs. (6)–(8) again, except that the terms involving the quantity of structures, $\beta^t(1-\delta A_n^t)S_n^t$ in each of the Eqs. (6)–(8), are now replaced by the terms $\beta^1 p^t(1-\delta A_n^t)(1+\gamma R_n^t)S_n^t$; β^1, δ, and γ are the parameters to be

[27]The correlation coefficient between the room variable R and the structure area S (not adjusted for net depreciation) in our data set is 0.4746, somewhat lower than we anticipated.

[28]Palmquist (1984, p. 397) is one such expert: "It would be anticipated that the number of square feet of living space would not simply have a linear effect on price. As the number of square feet increases, construction costs do not increase proportionally since such items as wall area do not typically increase proportionally. Appraisers have long known that price per square foot varies with the size of the house." The empirical results of Coulson (1992, p. 77) on this issue indicate a great deal of volatility in price, but for large structures, the price of structure per unit area trended up fairly strongly for his sample of U.S. properties.

estimated, p^t is Statistics Netherlands' New Dwelling Construction Cost Price Index for quarter t described in the Section 4, A_n^t is the age in decades of property n in quarter t, R_n^t is the number of rooms less 4 for property n in quarter t, and S_n^t is the area of structure n in quarter t. Note that A_n^t is equal to 0 if property n sold in quarter t is a new house and that R_n^t is equal to 0 if property n sold in quarter t has 4 rooms. In order to identify the parameters β^1, δ, and γ, we need the exogenous characteristics variables A_n^t and R_n^t to take on the value 0 for at least some observations (and the 0 values should not occur for exactly the same observations). Note also that if γ is equal to 0, then the present model reduces to Model 3 in the previous section.

Using the data for city "A" to estimate Model 4, the R^2 was 0.8736 as compared to 0.8723 for Model 3. The log likelihood was −16,222.6, an increase of 17.1 over Model 3 for the addition of only one parameter. Again, all parameters were significantly different from 0, including γ. The estimated decade depreciation rate was $\delta^* = 0.1089(0.00361)$. The first quarter values for the estimated three marginal land prices were $\alpha_S^{1^*} = 0.2207(0.0249)$, $\alpha_M^{1^*} = 0.3465(0.0560)$, and $\alpha_L^{1^*} = 0.1741(0.0307)$. The first quarter estimate for quality adjusted structures was $\beta^{1^*} = 1.0069(0.0212)$ or 1,006.9 Euros/m^2. This is an estimate of the building cost per meter squared in quarter 1 for a new house with four rooms, whereas in Model 3 we estimated the building costs for all houses, irrespective of the number of rooms (which ranged from 2 to 14). The estimated number of rooms parameter was $\gamma^{1^*} = 0.02759(0.00493)$. So in quarter 1, the increase in the price of a new structure per meter squared due to an additional room is estimated at 0.02759/1.0069 or +2.74%. This seems to be a reasonable quality premium.

As usual, Eqs. (9)–(15) were used to construct a chained Fisher index of land prices, denoted by P_{L4}. The estimated price in quarter t for a square meter of quality adjusted structures for a four room house is $\beta^{t^*} \equiv \beta^{1^*} p^t$, and we use this price series as our constant quality price series for structures. The corresponding constant quality quarter t quantity of structures is $S^{t^*} \equiv \sum_{n=1}^{N(t)} (1 - \delta^* A_n^t)(1 + \gamma^* R_n^t) S_n^t$.[29] These structures price and quantity series were combined with the three land price and quantity series to form a chained overall Fisher house price index P_4. The constant quality structures price index P_{S4} is a normalization of the series $\beta^{1^*}, \ldots, \beta^{22*}$. This index is identical to the structures price index

[29]Thus we are implicitly quality adjusting the quantities of houses with different room sizes into "standard" houses with four rooms using the quality adjustment factors $\gamma^* R_n^t$ for house n in quarter t. Thus we are forming a hedonic structures aggregate. Alternatively, instead of forming a quality adjusted aggregate, we could distinguish houses with differing number of rooms as separate types of housing and use index number theory to aggregate the 13 types of house into a structures aggregate. In this second interpretation, the quarter t structure price $\beta^{t^*} = \beta^{1^*} p^t$ applies to a new house with four rooms. The appropriate price (per m^2) for a new house with $5, 6, \ldots, 14$ rooms would be $\beta^{1^*} p^t (1 + \gamma *)$, $\beta^{1^*} p^t (1 + 2\gamma^*), \ldots, \beta^{1^*} p^t (1 + 10\gamma^*)$, and the price for a new house with two and three rooms would be $\beta^{1^*} p^t (1 - 2\gamma^*)$ and $\beta^{1^*} p^t (1 - \gamma^*)$. Thus in this second approach, we distinguish 13 types of house (according to their number of rooms) and calculate separate price and quantity series for all 13 types (adjusted for depreciation as well). However, if we then aggregate these series using Laspeyres, Paasche, or Fisher indexes, we would find that the resulting aggregate structures price index would be proportional to the $\beta^{1^*} p^t$ series. So the second method is actually equivalent to the first method.

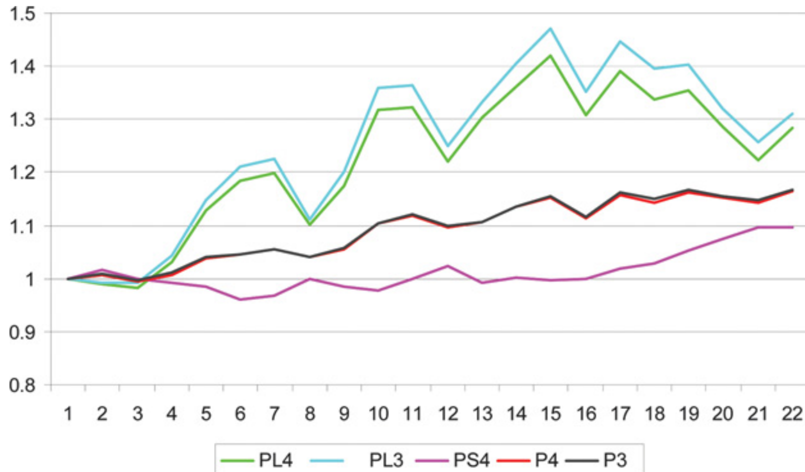

FIGURE 4 Land price indexes P_{L3} and P_{L4}, structures price index P_{S4}, and overall house price indexes P_3 and P_4.

P_{S3} for Model 3 because both models impose the same (nationwide) rates of change on quality adjusted structures prices.

The price indexes for land and structures and the overall house price index from Model 4 are listed in Table 2, along with the corresponding quantities. The three price indexes are plotted in Fig. 4; the previous price index for land, P_{L3}, and the previous overall price index, P_3, are also shown. Our new model generates a somewhat different series for the price of land as compared to Model 3; P_{L4} lies below P_{L3} for quarters 2–22. The overall house price indexes, P_3 and P_4, are virtually identical; i.e., they are difficult to distinguish in Fig. 4.[30]

Before running any regressions, we eliminated some outlier observations that had prices or characteristics which were either very large or very small relative to average prices and average amounts of characteristics. Observations with large error terms were not deleted, however. This nondeletion of regression outliers could affect our estimated coefficients, particularly if the outliers are either mostly positive or mostly negative. The empirical distribution of regression residuals appeared to be fairly symmetric with a

[30]If P_3 almost equals P_4 and P_{S3} is exactly equal to P_{S4}, one might ask how can P_{L3} and P_{L4} differ so much? The answer is that, while the rates of growth in the *price* of constant quality structures is the same in Models 3 and 4, the addition of the quality adjustment for the number of rooms has changed the initial level (and rates of growth) for the constant quality *quantity* of structures. Using Model 3, the initial levels of land and constant quality structures were 7,446.9 and 15,749.3. Using Model 4, the initial levels of land and constant quality structures were 8,372.8 and 14,816.2. Thus going from Model 3 to 4, the value of Q1 land has increased about 12.4%, and the value of structures has decreased to offset this increase. Since land prices increase more rapidly than structure prices and since the overall indexes P_3 and P_4 are virtually equal and the structures indexes P_{S3} and P_{S4} are exactly equal, it can be seen that these facts will imply that P_{L4} must grow more slowly than P_{L3}.

relatively small number of very large in magnitude residuals, and so these outliers do not seem to cause a major problem.

Our conclusion at this point is that Model 4 is a satisfactory hedonic housing regression model that enables us to decompose property prices into land and structures components. Also, the quality adjustments to the structures for age of the structure and for the number of rooms seem to be reasonable. The overall fit of the model to the data is satisfactory: an R^2 of 0.8736 for such a small number of characteristics is quite good. The Dutch data may of course not be representative of larger data sets, or data sets for other countries, where there might be more heterogeneity due to geography or differences in the types of houses being built over time.

Our builder's model could be further modified to account for additional characteristics. However, a certain amount of careful thought is required so that introducing additional characteristics reflect the realities of housing construction as well as locational effects. In particular, the number of stories in the dwelling unit is likely to be a significant quality adjustment characteristic: a higher number of stories (holding structural area constant) is likely to yield lower building costs due to shared floors and ceilings and less expenditures on roofing and insulation. A larger number of stories could also have a quality adjustment effect on the land component of the dwelling unit since a higher number of stories leads to more usable yard space. These construction realities will determine the appropriate functional form for the hedonic regression.

6. CONCLUSION

A number of tentative conclusions can be drawn from this study:

1. If we stratify housing sales by local area and type of housing and if we have data on the age of the dwelling unit, its land plot area (or share of the plot area in the case of multiple unit dwellings) and its floor space area, a wide variety of hedonic regression models that use these variables seem to yield much the same *overall* house price indexes.
2. It is much more difficult to obtain sensible land and structure price indexes by means of a hedonic regression. However, our builder's model, in conjunction with statistical agency information on the price changes of new dwellings, generated satisfactory results for our data set.
3. Adding the number of rooms in the dwelling unit as an explanatory variable in our hedonic regressions did improve the fit to the data but did not change the indexes substantially.
4. Splining land also improved the fit and led to a somewhat smoother land price series in our best builder's model.
5. It is important to delete observations in the regressions which are range outliers.

Some topics for follow up research include the following ones:

1. Can our method be generalized to deal with sales of condominiums and apartment units with shared land and facilities?
2. How exactly can other housing characteristics be used in more general versions of the builder's model?

ACKNOWLEDGMENTS

A preliminary version of this paper was presented at the Economic Measurement Group Workshop, 2009, December 9–11, Crowne Plaza Hotel, Coogee Beach, Sydney, Australia. The authors thank Christopher O'Donnell, Marc Francke, Ulrich Kohli, Alice Nakamura, Esmaralda Ramalho, Alicia Rambaldi, Mick Silver, Keith Woolford, and two referees for helpful comments. None of the individuals and institutions mentioned here and below are responsible for the contents of this paper.

FUNDING

The authors gratefully acknowledge the financial support from the Centre for Applied Economic Research at the University of New South Wales, the Australian Research Council (LP0347654 and LP0667655), and the Social Science and Humanities Research Council of Canada.

REFERENCES

Bostic, R. W., Longhofer, S. D., Readfearn, C. L. (2007). Land leverage: Decomposing home price dynamics. *Real Estate Economics* 35:183–2008.

Clapp, J. M. (1979). The substitution of urban land for other inputs. *Journal of Urban Economics* 6:122–134.

Clapp, J. M. (1980). The elasticity of substitution for land: The effects of measurement errors. *Journal of Urban Economics* 8:255–263.

Coulson, N. E. (1992). Semiparametric estimates of the marginal price of floorspace. *Journal of Real Estate Finance and Economics* 5:73–83.

Crone, T. M., Nakamura, L. I., Voith, R. P. (2009). Hedonic estimates of the cost of housing services: Rental and owner occupied units. In: Diewert, W. E., Balk, B. M., Fixler, D., Fox, K. J., Nakamura, A. O. (eds.), *Price and Productivity Measurement, Vol. 1: Housing*, Vancouver: Trafford Press, pp. 67–84.

Davis, M. A., Heathcote, J. (2007). The price and quantity of residential land in the United States. *Journal of Monetary Economics* 54:2595–2620.

Davis, M. A., Palumbo, M. G. (2008). The price of residential land in large U.S. cities. *Journal of Urban Economics* 63:352–384.

Diewert, W. E. (2007). The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and future directions. Discussion Paper 07-01, Department of Economics, University of British Columbia, Vancouver, British Columbia, Canada.

Diewert, W. E. (2010). Alternative approaches to measuring house price inflation. Discussion Paper 10-10, Department of Economics, University of British Columbia, Vancouver, British Columbia, Canada.

Diewert, W. E., Heravi, S., Silver, M. (2009). Hedonic imputation versus time dummy hedonic indexes. In: Diewert, W. E., Greenlees, J. S., Hulten, C. R. (eds.), *Price Index Concepts and Measurement, Studies in Income and Wealth*, Vol. 70. Chicago: University of Chicago Press, pp. 161–196.

Diewert, W. E., Nakamura, A. O., Nakamura, L. I. (2009). The housing bubble and a new approach to accounting for housing in a CPI. *Journal of Housing Economics* 18:156–171.

Diewert, W. E., de Haan, J., Hendriks, R. (2011). The decomposition of a house price index into land and structures: A hedonic regression approach. *The Valuation Journal* 6:58–105.

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

Fleming, M. C., Nellis, J. G. (1992). Development of standardized indices for measuring house price inflation incorporating physical and locational characteristics. *Applied Economics* 24:1067–1085.

Francke, M. K. (2008). The hierarchical trend model. In: Kauko, T., Damato, M. (eds.), *Mass Appraisal Methods: An International Perspective for Property Valuers*. Oxford: Wiley-Blackwell, pp. 164–180.

Francke, M. K. (2010). Repeat sales index for thin markets: A structural time series approach. *Journal of Real Estate Finance and Economics* 41:24–52.

Francke, M. K., Vos, G. A. (2004). The hierarchical trend model for property valuation and local price indices. *Journal of Real Estate Finance and Economics* 28:179–208.

Glaeser, E. L., Gyourko, J. (2003). The impact of building restrictions on housing affordability. *Economic Policy Review* 9:21–39.

Gouriéroux, C., Laferrère, A. (2009). Managing hedonic house price indexes: The French experience. *Journal of Housing Economics* 18:206–213.

Gyourko, J., Saiz, A. (2004). Reinvestment in the housing stock: The role of construction costs and the supply side. *Journal of Urban Economics* 55:238–256.

de Haan, J. (2009). Comment on hedonic imputation versus time dummy hedonic indexes. In: Diewert, W. E., Greenlees, J. S., Hulten, C. R. (eds.), *Price Index Concepts and Measurement, Studies in Income and Wealth*, Vol. 70. Chicago: University of Chicago Press, pp. 196–200.

de Haan, J. (2010). Hedonic price indexes: A comparison of imputation, time dummy and "Re-pricing" methods. *Journal of Economics and Statistics* (Jahrbücher für Nationalökonomie und Statistik) 230: 772–791.

Hill, R. J. (2011). Hedonic price indexes for housing. Statistics Directorate, Working Paper No 36, February 14, Paris: OECD.

Hill, R. J., Melser, D., Syed, I. (2009). Measuring a boom and bust: The Sydney housing market 2001–2006. *Journal of Housing Economics* 18:193–205.

Koev, E., Santos Silva, J. M. C. (2008). Hedonic methods for decomposing house price indices into land and structure components. Unpublished paper, Department of Economics, University of Essex, England.

McMillen, D. P. (2003). The return of centralization to Chicago: Using repeat sales to identify changes in house price distance gradients. *Regional Science and Urban Economics* 33:287–304.

Muth, R. F. (1971). The derived demand for urban residential land. *Urban Studies* 8:243–254.

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. *The Review of Economics and Statistics* 66:394–404.

Rambaldi, A. N., McAllister, R. R. J., Collins. K., Fletcher, C. S. (2010). Separating land from structure in property prices: A case study from Brisbane, Australia. School of Economics, University of Queensland, St. Lucia, Queensland, Australia.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82:34–55.

Rosen, H. S. (1978). Estimating inter-city differences in the price of housing services. *Urban Studies* 15: 351–355.

Schwann, G. M. (1998). A Real Estate Price Index for thin markets. *Journal of Real Estate Finance and Economics* 16:269–287.

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.