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

W. Erwin Diewert, Jan de Haan and Rens Hendriks,[1] April 9, 2011
Discussion Paper 11-01,
Department of Economics,
The University of British Columbia,
Vancouver, Canada, V6T 1Z1.
email: diewert@econ.ubc.ca

Abstract

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 Dutch city. In order to get sensible results, it was useful to use a nonlinear regression model using data that covered multiple time periods. It also proved to be necessary to impose some restrictions on the price of structures. The resulting builder's hedonic regression model was compared with the results for traditional logarithmic hedonic regression models.

Key Words

House price indexes, land and structure components, time dummy hedonic regressions, Fisher ideal indexes.

Journal of Economic Literature Classification Numbers

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

[1] 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. W. Erwin Diewert: Department of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and the School of Economics, University of New South Wales, Sydney, Australia (e-mail: diewert@econ.ubc.ca); Jan de Haan, Statistics Netherlands (email: Jhhn@cbs.nl) and Rens Hendriks, Statistics Netherlands (email: r.hendriks@cbs.nl). 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. 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. None of the above individuals and institutions are responsible for the contents of this paper.

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 to have a decomposition of housing values into land and structures components since structures depreciate while land does not. Thus our goal in this paper is to use readily available multiple listing data on sales of residential properties and to decompose the sales price of each property into a land component and a structures component. We will use the data pertaining to the sales of detached houses in a small Dutch city for 22 quarters, starting in Quarter 1 in 2003 and running to the end of Quarter 2 in 2008. We utilize a *hedonic regression approach* to accomplish our decomposition but our approach is based on a cost oriented model which we call the *builder's approach* to modeling hedonic regressions in the housing context. It is possible to use the more traditional time dummy approach to hedonic regression models in order to obtain a decomposition of property values into land and structures components but in section 9 below, we argue that the resulting valuations are not very realistic as compared to the valuations we obtain using the builder's approach. Another feature of our suggested approach is that it requires relatively little information on the characteristics of the houses that are in the data base: information on the plot area, the area of the structure, the age of the structure and the number of rooms in the house suffices to generate regression models that explain 85-89% of the variation in the selling prices of the houses in the data base.

A more detailed outline of the contents of this paper follows.

In section 2, we will consider a very simple hedonic regression model where we use information on only three characteristics of the property: the lot size, the size of the structure and the (approximate) age of the structure. We run a separate hedonic regression for each quarter which leads to estimated prices for land and structures for each quarter. These estimated characteristics prices can then be converted into land and structures prices covering the 22 quarters of data in our sample. We postulate 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. Thus our approach to the valuation of a residential property is essentially a crude cost of production approach. Note that the overall value of the property is assumed to be the *sum* of these two components.

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 2 break points. Again, a separate 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 very successful. The problem is due to multicollinearity and variability in the data and this volatility leads to a tendency for the regression models to fit the outliers, leading to erratic estimates for the price of land and structures.

We try two different methods in order to deal with the multicollinearity problem. In section 4, we impose *monotonicity restrictions* on the price of structures while in section 5, we draw on *exogenous information* on new house building costs from the national statistical agency and assume that the price movements for new structures mirror the statistical agency movements in the price of new houses. We find that the use of exogenous information generates the most reasonable decomposition of house values into their structure and land components.

In section 6, we generalize the model in section 5 to include information on the number of rooms in the house as an additional price determining characteristic. The idea here is that a higher number of rooms in a house generally indicates that the quality of construction of the house will be higher. Our regression results support this hypothesis: the estimated increase in the price of a new structure per m^2 in Quarter 1 due to an additional room is about 2.7%.

Sections 7-9 estimate various time dummy hedonic regression models using the same data set so that the results for these more traditional hedonic housing regressions models can be contrasted with the results for our builder's model. In section 7, selling price is the dependent variable and the characteristics are entered without any transformation; i.e., in linear form. In section 8, log selling price is the dependent variable and the characteristics are entered in linear form. Finally in section 9, we estimate a traditional log-log time dummy model. This model generates an overall house price index that is close to the overall house price index generated by our builder's model but the structure and land components associated with this index are not very realistic. The conclusion that we draw from sections 7-9 is that traditional time dummy hedonic regression models for housing can generate very reasonable *overall* house price indexes but they cannot generate very reasonable component house price indexes.

We conclude this section by providing a brief literature review of methods used to provide a decomposition of the selling price of a dwelling unit into land and structures components. Basically, variations of *three methods* have been used:

- The vacant land method;
- The construction cost method and
- The hedonic regression method.

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 $p_L L$:

(1) $V = p_L L + p_S S$

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 equation (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; 125) (1980; 256) and he noted that the method 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 then this price is applied to the comparable properties and equation (1) can then be used to solve for the structure price p_S. This method was used by Thorsnes (1997) and Bostic, Longhofer and Readfearn (2007).[2]

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, equation (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) where 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.[3]

A variant of the *hedonic regression method* is the method that will be used in this paper. Various versions of the method will be explained in sections 2-6 and 9 below. Some early papers that use a similar methodology include Clapp (1980), Palmquist (1984), Fleming and Nellis (1992) and Schwann (1998). Basically, 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 and these marginal prices can be used to decompose the house value into land and structures components under certain conditions.

2. Model 1: A Simple Builder's Model

[2] The set of vacant lots can be augmented by properties which are sold and the associated structure is immediately demolished. Clapp (1980; 256) lists several reasons why the vacant land method is not likely to be very accurate.

[3] Muth (1971; 246) and Rosen (1978; 353-354) 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 equation (1), S was determined residually.

Hedonic regression models are frequently used to obtain constant quality price indexes for owner occupied housing.[4] Although there are many variants of the technique, the basic model regresses the logarithm of the sale price of the property on the price determining characteristics of the property and a time dummy variable is added for each period in the regression (except the base period). Once the estimation has been completed, these time dummy coefficients can be exponentiated and turned into an index.[5]

A residential property has a number of important price determining characteristics:

- The land area of the property (L);
- The livable floor space area of the structure (S);
- The age of the structure (A);
- The type of dwelling unit (detached, row, apartment);
- The type of construction (wood, brick, concrete);
- The location of the property.[6]

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 variable since the house 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 A, structure size in meters squared S and on the age A of the structure. We will find that hedonic regression models that use only these three explanatory variables give rise to an R^2 that is in the range .87 to .88, which indicates that most of the variation in the data can be explained by using just these three variables.[7]

As noted in the introduction, for some purposes, it would be very useful to decompose the overall price of a property into *additive components* that reflected the value of the land that the structure sits on and the value of the structure. The primary purpose of the present paper is to determine whether a hedonic regression technique could provide such a decomposition.

---

[4] For some recent literature, see Crone, Nakamura and Voith (2009), Diewert, Nakamura and Nakamura (2009), Gouriéroux and Laferrère (2009), Hill (2011), Hill, Melser and Syed (2009) and Hill (2011).

[5] An alternative approach to the time dummy hedonic method is to estimate separate hedonic regressions for both of the periods compared; this is called the hedonic imputation approach. See Haan (2008) (2009) and Diewert, Heravi and Silver (2009) for theoretical discussions and comparisons between these alternative approaches.

[6] There are many other price determining characteristics that could be added to this list such as landscaping, the number of floors and rooms, 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 but if the location of the properties in the sample of sales is small enough, then it should not be necessary to add these characteristics. In our example, the Dutch town of “A” is small enough and homogeneous enough so that these neighbourhood effects can be neglected. In other cities or neighborhoods where geography creates important locational differences, our rather minimal basic model will probably not fit the data as well. Our simple builder’s model will probably not work well for multiple unit structures where the height of the apartment becomes an important price determining characteristic.

[7] In sections 6-9, we add the number of rooms as an additional explanatory variable.

Several researchers have suggested hedonic regression models that lead to *additive decompositions* of an overall property price into land and structures components.[8] We will now outline Diewert's (2007) justification for an additive decomposition.

If we momentarily think like 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 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 V_n^t in period t[9] and structure areas S_n^t and land areas L_n^t for n = 1,...,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.[10] This leads to the following hedonic regression model for period t where α^t and β^t are the parameters to be estimated in the regression:[11]

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

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 (1) for each period t in our sample.

The hedonic regression model defined by (1) is the simplest possible one but it applies only to new structures. But it is likely that a model that is similar to (1) applies to older structures as well. Older structures will be worth less than newer structures due to the depreciation (or deterioration due to aging effects) of the structure. Thus 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

[8] See Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic, Longhofer and Redfearn (2007), Davis and Heathcote (2007), Diewert (2007), Francke (2008), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert, Haan and Hendriks (2010) and Diewert (2010).
[9] Note that we have labeled these property prices as V_n^0 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.
[10] 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.
[11] In order to obtain homoskedastic errors, it would be preferable to assume multiplicative errors in equation (1) 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 (1) since we are attempting to decompose the aggregate value of housing (in the sample of properties that sold during the period) into additive structures and land components and the additive error specification will facilitate this decomposition.

more realistic hedonic regression model than that defined by (1) above is the following *basic builder's model*:[12]

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

where the parameter δ^t reflects the *net depreciation rate* as the structure ages one additional period. Thus if the age of the structure is measured in years, we would expect δ^t to be between 1 and 2%.[13] Note that (2) is now a nonlinear regression model whereas (1) was a simple linear regression model.[14] Both models (1) and (2) can be run period by period; it is not necessary to run one big regression covering all time periods in the data sample. The period t price of land will the estimated coefficient for the parameter α^t and the price of a unit of a newly built structure for period t will be the estimate for β^t. The period t quantity of land for property n is $L_n{}^t$ 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 implemented the above model (2) using real estate sales 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 Q1 2003 and extending through Q2 in 2008 (so our T = 22). The data that we used can be described as follows:

- $V_n{}^t$ is the selling price of property n in quarter t in units of 1,000 Euros where t = 1,...,22;
- $L_n{}^t$ is the area of the plot for the sale of property n in quarter t in units of meters squared;[15]
- $S_n{}^t$ is the living space area of the structure for the sale of property n in quarter t in units of meters squared;
- $A_n{}^t$ is the (approximate) age in decades of the structure on property n in period t;[16]

[12] Note that the model in this section is a *supply side model* as opposed to the *demand side model* of Muth (1971) and McMillen (2003) which will be explained in section 9 below. Basically, we are assuming identical 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 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.

[13] 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.

[14] This formulation follows that of Diewert (2007) and Diewert, Haan and Hendriks (2010). It is a special case of Clapp's (1980; 258) model except that Clapp included a constant term.

[15] 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 was Shazam (the nonlinear option).

- 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 in sections 3-5, we did not use the room variable R as an explanatory variable.[17]

Initially, there were 3543 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. Thus we looked at the range of our V, L, S and R variables and deleted 54 range outliers. There were also two duplicate observations in Q1 for 2006 and these duplicates were also deleted. Thus we ended up with 3487 data points for the 22 quarters.[18] The sample means for the data with outliers excluded (standard deviations in brackets) were as follows: $\overline{V}$ = 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). Thus the entire sample of houses sold at the average price of 182,260 Euros, the average plot size was 258.1 m^2, the average living space in the structure was 126.6 m^2 and the average age was approximately 18.9 years. The sample median price was 160,000 Euros.

The correlations between the various variables are also of interest. The correlation coefficients of the selling price V with L, S, A and R are .8014, .7919, −.3752 and .3790 respectively.[19] Thus the selling price V is fairly highly correlated with both land L and (unadjusted) structures S. The correlation between L and S is .6248 and thus there is the possibility of multicollinearity between these variables. Finally there is also a substantial positive correlation of .4746 between the structure area S and the number of rooms R.

Instead of running 22 quarterly regressions of the form (2), we combined the data using dummy variables and ran one big regression, which combined all 22 quarterly regressions into a single regression.[20] The R^2 for the resulting combined regression was .8729, which is quite good, considering we have only 3 explanatory variables (but 66 parameters to estimate). The resulting log likelihood was −16231.6. The results of our 22 nonlinear

---

[16] 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, $A_n^t = 0$.

[17] In sections 6-9 below, we did use the room variable as a quality adjustment variable.

[18] 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 1500 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.

[19] In order to illustrate the importance of deleting range outliers for all variables, the correlation coefficients of V with L, S, A and R for the original data set with 3543 observations was 0.33331, 0.80795, −0.34111 and 0.34291. Thus it is particularly important to delete land outliers.

[20] This one 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.

regressions of the type defined by (2) above are summarized in Table 1 below.[21] The *quality adjusted structures quantity in quarter t*, S^{t*}, is equal to the sum over the properties sold n in that quarter adjusted into new structure units; i.e., $S^{t*} \equiv \sum_{n\in N(t)} (1 - \delta^{t*}A_n^{\,t})S_n^{\,t}$.

Table 1: Estimated Land Prices α^{t*}, Structure Prices β^{t*}, Decade Depreciation Rates δ^{t*}, Land Quantities L^t and Quality Adjusted Structures Quantities S^{t*}

Quarter	α^{t*}	β^{t*}	δ^{t*}	L^t	S^{t*}
1	0.24244	0.96475	0.08795	35023	15185.0
2	0.30516	0.88095	0.11366	35412	13642.4
3	0.18981	1.02192	0.06383	39872	16101.0
4	0.26257	0.97406	0.10083	42449	16828.6
5	0.28735	0.94982	0.10094	37319	14836.4
6	0.24126	1.09852	0.10294	45611	16807.6
7	0.29896	1.03848	0.13716	33321	12165.4
8	0.23997	1.10097	0.11281	40395	16799.6
9	0.21462	1.10121	0.07683	38578	17556.5
10	0.29719	1.00176	0.08451	38246	16455.1
11	0.26206	1.13478	0.09354	39112	16131.8
12	0.21629	1.18213	0.08516	41288	16688.1
13	0.28200	1.09846	0.11428	43387	16425.0
14	0.24724	1.20922	0.10188	46132	19058.1
15	0.31203	1.11738	0.11721	39250	15361.6
16	0.28787	1.15708	0.13372	40102	14767.3
17	0.29934	1.08880	0.08949	39813	16703.1
18	0.24951	1.29435	0.12144	56992	19664.0
19	0.33990	1.12320	0.13708	35801	12476.6
20	0.26649	1.19169	0.09642	48031	19861.5
21	0.20590	1.28938	0.08753	37854	15839.4
22	0.34341	1.02286	0.09684	45878	19858.6

It can be seen that the decade net depreciation rates δ^{t*} are 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 rate. Unfortunately, our estimated land and structures prices are not at all reasonable: the price of land sinks to a very low level in quarter 3 while the price of structures has a local peak in this quarter. In general, the land and constant quality structures prices are too volatile.[22]

It seems reasonable to assume a net depreciation rate that is constant across quarters and so the model defined by (2) is replaced by the following *Model 1*:[23]

[21] We used the nonlinear option in Shazam to run our regressions. The lowest t statistic for the parameters estimated in this model was 5.8 for the Q1 estimated depreciation rate δ^{1*}.
[22] This period to period parameter instability problem was noted by Schwann (1998; 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."
[23] We regard the previous model as a preliminary model, Model 0.

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

where the parameter δ reflects the sample *net depreciation rate* as the structure ages one additional decade but now it is assumed to be constant over the entire sample period. Thus 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 equations (2).

The R^2 for the resulting nonlinear regression model was .8703,[24] which is quite good, considering we have only 2 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 .8729. The log likelihood for the constant depreciation rate model was −16266.6, which is a decrease of 35.0 from the log likelihood of the previous model. This decrease in log likelihood seems to be a reasonable price to pay in order to obtain a stable estimate for the net depreciation rate. Note that the estimated decade net depreciation rate is now $\delta^* = 0.10241$ or about 1% per year. The smallest t statistic for the parameters in this model was 11.9 for the parameter α^{1*}. The results for our new model (3) are summarized in Table 2 below. The estimated *quality adjusted structures quantity in quarter t*, S^{t*}, is equal to the sum over the properties sold n in that quarter, quality adjusted (for net depreciation) into new structure units; i.e.:

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

where δ^* is the estimated net depreciation rate for the entire sample period.

Table 2: Estimated Land Prices α^{t*}, Structure Prices β^{t*}, the Decade Depreciation Rate δ^*, Land Quantities L^t and Quality Adjusted Structures Quantities S^{t*}

Quarter	α^{t*}	β^{t*}	δ^*	L^t	S^{t*}
1	0.25162	0.97205	0.10241	35023	14677.2
2	0.30084	0.86961	0.10241	35412	14047.9
3	0.20130	1.07050	0.10241	39872	14680.1
4	0.26348	0.97486	0.10241	42449	16764.0
5	0.28792	0.95083	0.10241	37319	14787.8
6	0.24087	1.09845	0.10241	45611	16828.1
7	0.27564	1.02882	0.10241	33321	13234.3
8	0.23536	1.09186	0.10241	40395	17169.1
9	0.23548	1.10259	0.10241	38578	16680.0
10	0.30717	1.00917	0.10241	38246	15847.6
11	0.26523	1.14512	0.10241	39112	15831.3
12	0.22357	1.19693	0.10241	41288	16119.8
13	0.27415	1.09353	0.10241	43387	16873.5
14	0.24764	1.20932	0.10241	46132	19037.4
15	0.30056	1.11530	0.10241	39250	15889.7
16	0.26941	1.13981	0.10241	40102	15836.9
17	0.31121	1.08539	0.10241	39813	16234.7

[24] 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.

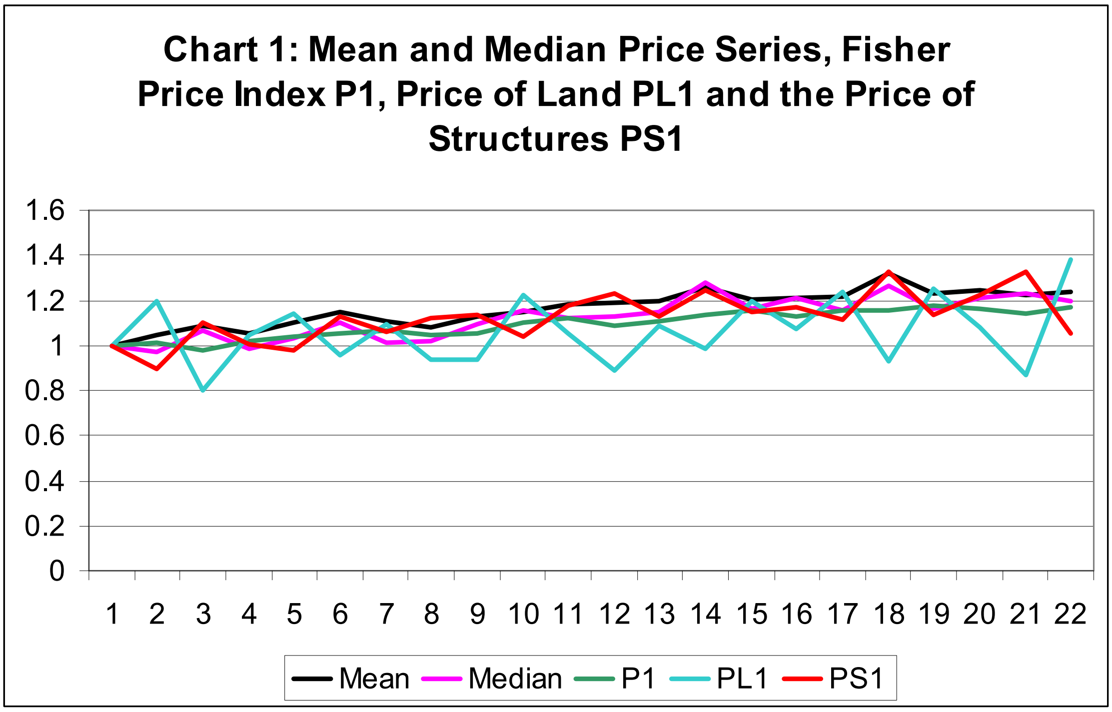
18	0.23368	1.28996	0.10241	56992	20579.3
19	0.31558	1.10402	0.10241	35801	13661.4
20	0.27131	1.19228	0.10241	48031	19610.7
21	0.21835	1.29223	0.10241	37854	15344.4
22	0.34704	1.02324	0.10241	45878	19645.7

It is of some interest to compare the above land and structures prices with the mean and median prices for houses in the sample for each quarter. These prices were normalized to equal 1 in quarter 1 and are listed as P_{Mean} and P_{Median} in Table 3 below. The land and structures prices in Table 1, α^{t*} and β^{t*}, were also normalized to equal 1 in quarter 1 and are listed as P_{L1} and P_{S1} in Table 3. Finally, we used the price data in Table 2, α^{t*} and β^{t*}, along with the corresponding quantity data, L^t and S^{t*}, in Table 1 in order to calculate a "constant quality" chained Fisher (1922) house price index, which is listed as P_1 in Table 3.

Table 3: Quarterly Mean, Median and Fisher Housing Prices P_1 and the Price of Land P_{L1} and Structures P_{S1}

Quarter	P_{Mean}	P_{Median}	P_1	P_{L1}	P_{S1}
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.04916	0.97007	1.01150	1.19559	0.89461
3	1.08473	1.06796	0.97511	0.80001	1.10128
4	1.05544	0.98592	1.01626	1.04711	1.00289
5	1.10128	1.03521	1.03964	1.14425	0.97817
6	1.14688	1.10035	1.05462	0.95727	1.13004
7	1.10436	1.01408	1.06757	1.09546	1.05840
8	1.07874	1.02113	1.04559	0.93537	1.12326
9	1.12774	1.09155	1.05259	0.93584	1.13429
10	1.15032	1.15493	1.10079	1.22074	1.03819
11	1.18601	1.12148	1.12179	1.05409	1.17805
12	1.19096	1.12676	1.08897	0.88850	1.23134
13	1.19633	1.14789	1.10521	1.08951	1.12497
14	1.26120	1.28169	1.13606	0.98418	1.24409
15	1.20159	1.16197	1.15825	1.19450	1.14737
16	1.21170	1.21303	1.12513	1.07071	1.17258
17	1.21731	1.15493	1.15603	1.23682	1.11660
18	1.31762	1.26761	1.15751	0.92870	1.32705
19	1.22870	1.16056	1.17844	1.25419	1.13576
20	1.24592	1.20775	1.16364	1.07825	1.22656
21	1.22596	1.23239	1.14472	0.86778	1.32939
22	1.23604	1.19718	1.16987	1.37920	1.05266
Mean	1.1645	1.1234	1.0941	1.0617	1.1249

It can be seen that the mean and median series are rather volatile and differ substantially from P_1, the Fisher index that is compiled using the results of our builder's regression model (3) using the data on the price of land P_{L1} and quality adjusted structures P_{S1} and the associated quantities tabled in Table 2 above. The overall Fisher house price index P_1 is fairly smooth but its component prices P_{L1} and P_{S1} fluctuate violently. The price series listed in Table 2 are graphed in Chart 1.



It can be seen that the Mean and Median price series are on average substantially above the corresponding overall Fisher house price index P_1 and the series P_1 is much smoother.[25] It appears that the P_1 series provides satisfactory estimates for the overall price of houses. On the other hand, the component land and structure price series for P_1, P_{L1} and P_{S1}, are extremely volatile and hence are not very credible estimates for the underlying movements for the price of land and constant quality structures in the town of "A" over this period. It can be seen that when the price of land spikes up, the corresponding price of structures tends to spike downwards and vice versa. This erratic behavior in P_{L1} and P_{S1} is due to measurement errors in the quantity of land and the quantity of structures[26] along with a substantial correlation between the quantity of land and structures; i.e., we have a multicollinearity problem.

One possible problem with our highly simplified house price model is that our model makes no allowance for the fact that larger sized plots tend to sell for an average price that is below the price for medium and smaller sized plots. Thus in the following section, we will generalize the builder's model (3) to take into account this empirical regularity.

3. Model 2: The Use of Linear Splines on Lot Size

[25] We attribute the slower rate of growth in our hedonic index P_1 as compared to the Mean and Median indexes to the fact that new houses tend to get bigger over time. The Mean and Median indexes cannot take this quality improvement into account.

[26] The measurement errors here are include recording errors but also include errors due to our imperfect measurement of the quality of construction and the quality of the land; e.g., 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 in fact, this will not be true.

In most countries, the reality is that large lots tend to sell at a lower price per unit area than smaller lots.[27] Thus in this section, we will assume that builders face a piecewise linear schedule of prices per unit land when they purchase a lot. This linear spline model will allow the price of large lots to drop as compared to smaller lots. We broke up our 3487 observations into 3 groups of property sales:

- 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).

The small lot size group had 1194 sales, the medium lot size group 1108 sales and the large lot size group had 1185 sales, so that the three groups were roughly equal in size. We define the sets of observations n which belong to Group S, M and L in period t to be $N_S(t)$, $N_M(t)$ 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 in (3) above; i.e., the following estimating equation was used:

$$(5)\ V_n{}^t = \alpha_S{}^t L_n{}^t + \beta^t(1 - \delta A_n{}^t)S_n{}^t + \varepsilon_n{}^t\ ; \qquad t = 1,...,22;\ n \in N_S(t)$$

where the unknown parameters to be estimated are $\alpha_S{}^t$, β^t for t = 1,...,22 and δ. For an observation n in period t that was associated with a medium lot size, the following estimating equation was used:

$$(6)\ V_n{}^t = \alpha_S{}^t(170) + \alpha_M{}^t(L_n{}^t - 170) + \beta^t(1 - \delta A_n{}^t)S_n{}^t + \varepsilon_n{}^t\ ; \qquad t = 1,...,22;\ n \in N_M(t)$$

where we have added 22 new parameters to be estimated, the $\alpha_M{}^t$ for t = 1,...,22. Finally, for an observation n in period t that was associated with a large lot size, the following estimating equation was used:

$$(7)\ V_n{}^t = \alpha_S{}^t(170) + \alpha_M{}^t(270 - 170) + \alpha_L{}^t(L_n{}^t - 270) + \beta^t(1 - \delta A_n{}^t)S_n{}^t + \varepsilon_n{}^t\ ; \qquad t = 1,...,T;\ n \in N_L(t)$$

where we have added 22 new parameters to be estimated, the $\alpha_L{}^t$ for t = 1,...,22. Thus for small lots, the value of an extra marginal addition of land in quarter t is $\alpha_S{}^t$, for medium lots, the value of an extra marginal addition of land in quarter t is $\alpha_M{}^t$ and for large lots, the value of an extra marginal addition of land in quarter t is $\alpha_L{}^t$. These pricing schedules

[27] This empirical regularity was noted by Francke (2008; 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, as shown in Figure 8.1. The first 300 m^2 counts for 100%, from 300 m^2 until 500 m^2 counts for 53% and so on." At first glance, it appears that Francke is using a step function to model the price schedule but in fact, he used linear splines in the same way as the present authors.

are joined together so that the cost of an extra unit of land increases with the size of the lot in a continuous fashion.[28] The above model can readily be put into a nonlinear regression format for each period using dummy variables to indicate whether an observation is in Group S, M or L. The nonlinear option in Shazam was used to estimate *Model 2* defined by (5)-(7) as one big regression.

The R^2 for this model was .8756, a substantial increase over the previous two models (without splines) where the R^2 was .8729 (many depreciation rates) and .8703 (one depreciation rate). The new log likelihood was −16195.0, an increase of 71.6 from the previous model's log likelihood. The estimated decade depreciation rate was $\delta^* = 0.1019$ (0.00329).[29] The first period parameter values for the 3 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 of small lots was estimated to be 288.9 Euros per m^2. For medium sized lots, the estimated marginal cost was 364.3 Euros/m.^{2} For large lots, the estimated marginal cost was 189.5 Euros/m^2. 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/m^2. All of the estimated coefficients were positive. The lowest t statistic for all of the 89 parameters was 2.79 (for α_S^8), so all of the estimated coefficients in this model were significantly different from zero. Our conclusion is that adding splines for the lot size gives us additional explanatory power.

Once the parameters for the model have been estimated, then in each quarter t, we can calculate the *predicted value of land for small, medium and large lot sales*, V_{LS}^t, V_{LM}^t and V_{LL}^t respectively, along with the associated *quantities of land*, L_{LS}^t, L_{LM}^t and L_{LL}^t as follows:

$$(8)\quad V_{LS}^t \equiv \sum_{n\in N_S(t)} \alpha_S^{t*} L_n^t \; ; \qquad t = 1,\ldots,22;$$

$$(9)\quad V_{LM}^t \equiv \sum_{n\in N_M(t)} \alpha_S^{t*}[170] + \alpha_M^{t*}[L_n^t - 170] \; ; \qquad t = 1,\ldots,22;$$

$$(10)\quad V_{LL}^t \equiv \sum_{n\in N_L(t)} \alpha_S^{t*}[170] + \alpha_M^{t*}[100] + \alpha_L^{t*}[L_n^t - 270] \; ; \qquad t = 1,\ldots,22;$$

$$(11)\quad L_{LS}^t \equiv \sum_{n\in N_S(t)} L_n^t \; ; \qquad t = 1,\ldots,22;$$

$$(12)\quad L_{LM}^t \equiv \sum_{n\in N_M(t)} L_n^t \; ; \qquad t = 1,\ldots,22;$$

$$(13)\quad L_{LL}^t \equiv \sum_{n\in N_L(t)} L_n^t \qquad t = 1,\ldots,22.$$

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

$$(14)\quad P_{LS}^t \equiv V_{LS}^t/L_{LS}^t \; ; \; P_{LM}^t \equiv V_{LM}^t/L_{LM}^t \; ; \; P_{LL}^t \equiv V_{LL}^t/L_{LL}^t \; ; \qquad t = 1,\ldots,22.$$

[28] Thus 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.

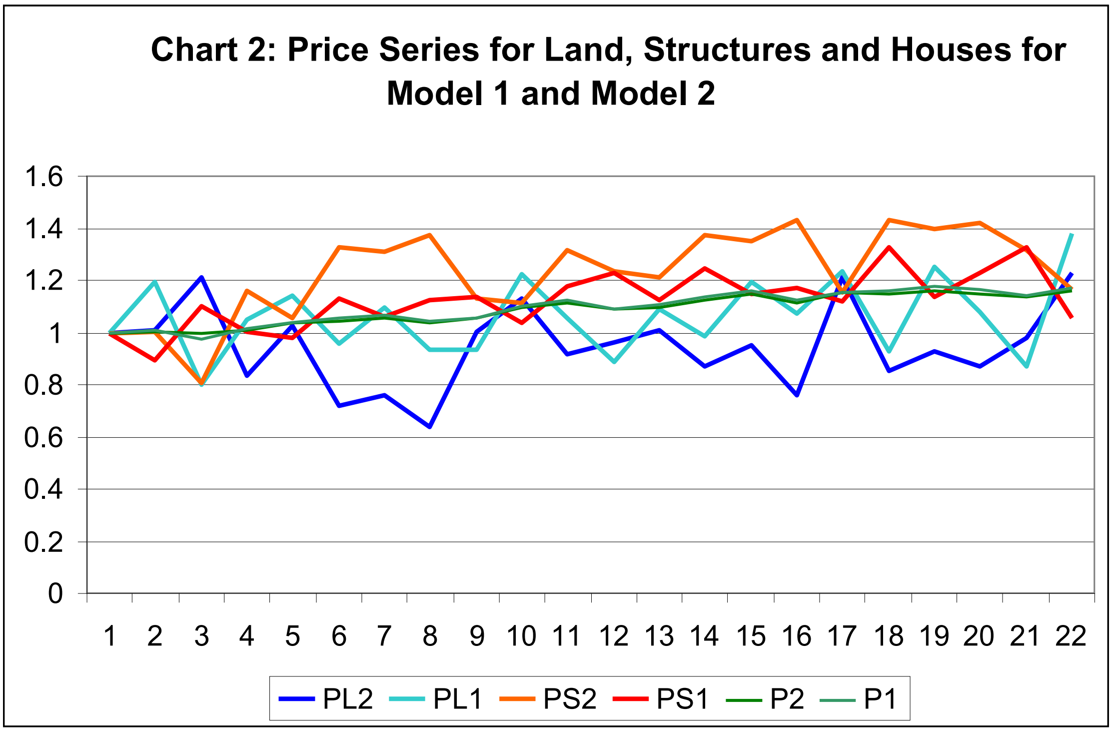
[29] Standard errors are in brackets.

The average land prices for small, medium and large lots defined by (14) and the corresponding quantities of land defined by (11)-(13) can be used to form a chained *Fisher land price index*, which we denote by P_{L2}. This index is plotted in Chart 2 and listed in Table 4 below. As in the previous model, the estimated period t price for a square meter of quality adjusted structures 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 structures 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 which is graphed in Chart 2 and listed in Table 4. The *constant quality structures price index* P_{S2} (a normalization of the series $\beta^{1*},...,\beta^{22*}$) is also found in Chart 2 and Table 4.

Table 4: The Price of Land P_{L2}, the Price of Structures P_{S2} and the Overall House Price Index P_2 Generated by Model 2 with the Corresponding Quantities Q_{L2}, Q_{S2} and Q_2

Quarter	P_{L2}	P_{S2}	P_2	Q_{L2}	Q_{S2}	Q_2
1	1.00000	1.00000	1.00000	10206.6	12974.2	23180.8
2	1.00908	1.00275	1.00560	10301.6	12419.0	22721.1
3	1.21368	0.80574	0.99438	11399.0	12977.5	24428.4
4	0.83401	1.16101	1.01036	12244.5	14819.2	27136.1
5	1.02424	1.05780	1.03968	10761.0	13070.8	23899.8
6	0.72144	1.32784	1.04623	13249.3	14874.9	28014.9
7	0.75915	1.30870	1.05326	9559.8	11698.3	21425.8
8	0.63613	1.37463	1.03651	11325.6	15174.5	27075.3
9	1.00299	1.13143	1.05271	10847.9	14742.2	26180.1
10	1.13249	1.11450	1.09760	10748.2	14007.1	25312.6
11	0.91532	1.31447	1.11485	10987.1	13992.6	25518.8
12	0.96261	1.23413	1.09171	11588.5	14247.0	26323.6
13	1.00878	1.20893	1.09851	12237.9	14914.5	27652.1
14	0.86737	1.37357	1.12687	13026.3	16825.5	30535.7
15	0.94969	1.35088	1.14963	10978.1	14045.0	25572.5
16	0.76142	1.43021	1.11129	11351.7	13997.7	25792.7
17	1.21221	1.15207	1.15468	11215.6	14349.8	26091.8
18	0.85042	1.43432	1.14520	16126.5	18190.9	34758.8
19	0.92861	1.39930	1.16228	10043.1	12076.9	22563.7
20	0.86731	1.41818	1.14581	13473.0	17333.0	31651.6
21	0.97973	1.31754	1.13869	10607.8	13562.4	24819.6
22	1.22862	1.16373	1.15847	12866.2	17362.3	31086.4
Mean	0.94842	1.2310	1.0879	11598	14439	26433

In the following Chart, we will compare the price series P_{L2}, P_{S2} and P_2 generated by Model 2 with the price series P_{L1}, P_{S1} and P_1 that were generated by Model 1 in the previous section (which did not include splines on the size of the land area).



It can be seen that again there is a volatility problem with the price of land P_{L2} and the price of structures P_{S2} in our new builder's model with splines on land: when the price of land jumps up, the price of structures drops down and in fact, the offsetting jumps are now bigger than they were using the no splines model with a constant depreciation rate that was described at the end of the previous section. This offsetting volatility is again an indication of a severe multicollinearity problem. However, note that both models generate essentially the same overall house price index, which is quite smooth and looks reasonable; i.e., P_1 and P_2 can hardly be distinguished in Chart 2.

Due to the high correlation between the size of the structure and the size of the underlying plot and the measurement error in our land and quality adjusted structures series, it is going to be a difficult task to extract meaningful price and structure components out of information on house sales alone. Thus in the following two sections, we will add some *additional restrictions* on our basic model described in this section in attempts to obtain more meaningful land and structures price series.[30]

4. Model 3: The Use of Monotonicity Restrictions on the Price of Structures

[30] Another approach to the volatility problem is to use a *smoothing method* in order to stabilize the volatile period to period characteristics prices. This approach dates back to Coulson (1992) and Schwann (1998) and more recent contributions include Francke and Vos (2004), Francke (2009) and Rambaldi, McAllister, Collins and Fletcher (2011). We have not pursued this approach because we feel that it is not an appropriate one for statistical agencies who have to produce non-revisable housing price indexes in real time. The use of smoothing methods is appropriate when the task is to produce historical series but smoothing methods do not work well in a real time context due to the inability of these methods to predict turning points in the series.

A priori, it seemed likely that Dutch construction costs did not fall significantly during the sample period.[31] If this is the case, then monotonicity restrictions on the quarterly prices of quality adjusted structures, β^1, β^2, β^3,..., β^{22}, could be imposed on the hedonic regression model (5)-(7) without much loss of fit by replacing the constant quality quarter t structures price parameters β^t by the following sequence of parameters for the 22 quarters: β^1, $\beta^1 + (\phi^2)^2$, $\beta^1 + (\phi^2)^2 + (\phi^3)^2$,..., $\beta^1 + (\phi^2)^2 + (\phi^3)^2 + ... + (\phi^{14})^2$ where ϕ^2, ϕ^3,..., ϕ^{22} are scalar parameters.[32] Thus for each quarter t starting at quarter 2, the price of a square meter of constant quality structures γ^t is equal to the previous period's price γ^{t-1} plus the square of a parameter ϕ^{t-1}, $[\phi^{t-1}]^2$, for t = 2, 3,..., 14. Now replace this reparameterization of the structures price parameters β^t in equations (5)-(7) in order to obtain a *linear spline model for the price of land with monotonicity restrictions on the price of constant quality structures*. The resulting regression model is now considerably more nonlinear than the previous model but convergence did not prove to be a problem.

Using our 22 quarters of data for the town of "A", the estimated decade depreciation rate was $\delta^* = 0.1008$ (0.00324). The R^2 for this model was .8745, a drop from the previous unrestricted spline Model 2 where the R^2 was .8756. The log likelihood was −16209.8, a decrease of 14.8 over the previous unrestricted model. Sixteen of the 21 new parameters ϕ^t were zero in this monotonicity restricted regression. The first period parameter values for the 3 marginal prices for land were $\alpha_S^{1*} = 0.32530$ (0.0310), $\alpha_M^{1*} = 0.3700$ (0.0588) and $\alpha_L^{1*} = 0.1961$ (0.0308). Thus in quarter 1, the marginal cost per m^2 of small lots was estimated to be 325.3 Euros per m^2.[33] For medium sized lots, the estimated marginal cost was 370.0 Euros/m^2. For large lots, the estimated marginal cost was 196.1 Euros/m^2. The first period parameter value for quality adjusted structures was $\beta^{1*} = 0.8159$ (0.0402) so that a square meter of new structure was valued at 815.9 Euros/m^2.

Once the parameters for the model have been estimated, then the estimated ϕ^{t*} parameters were converted into β^{t*} parameters using the following recursive equations:

$$(15)\quad \beta^{t+1*} \equiv \beta^{t*} + [\phi^{t*}]^2\,; \qquad t = 2,...,14.$$

Now use equations (8)-(14) in the previous section in order to construct a chained Fisher index of land prices, which we denote by P_{L3}. This index is plotted in Chart 3 and listed in Table 5 below. As in the previous model, the estimated period t price for a square meter of quality adjusted structures 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 structures 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 which is graphed in Chart 3 and listed in Table 5. The

[31] Some direct evidence on this assertion will be presented in the following section. This direct evidence indicates that there were some fairly substantial movements in construction prices during this period.

[32] This method for imposing monotonicity restrictions was used by Diewert, Haan and Hendriks (2010) with the difference that they imposed monotonicity on both structures and land prices, whereas here, we impose monotonicity restrictions on structure prices only.

[33] Recall that we rescaled the original prices V_n^t by dividing them by 1000.

constant quality structures price index P_{S3} (a normalization of the series $\beta^{1*},...,\beta^{22*}$) is also found in Chart 3 and Table 5.

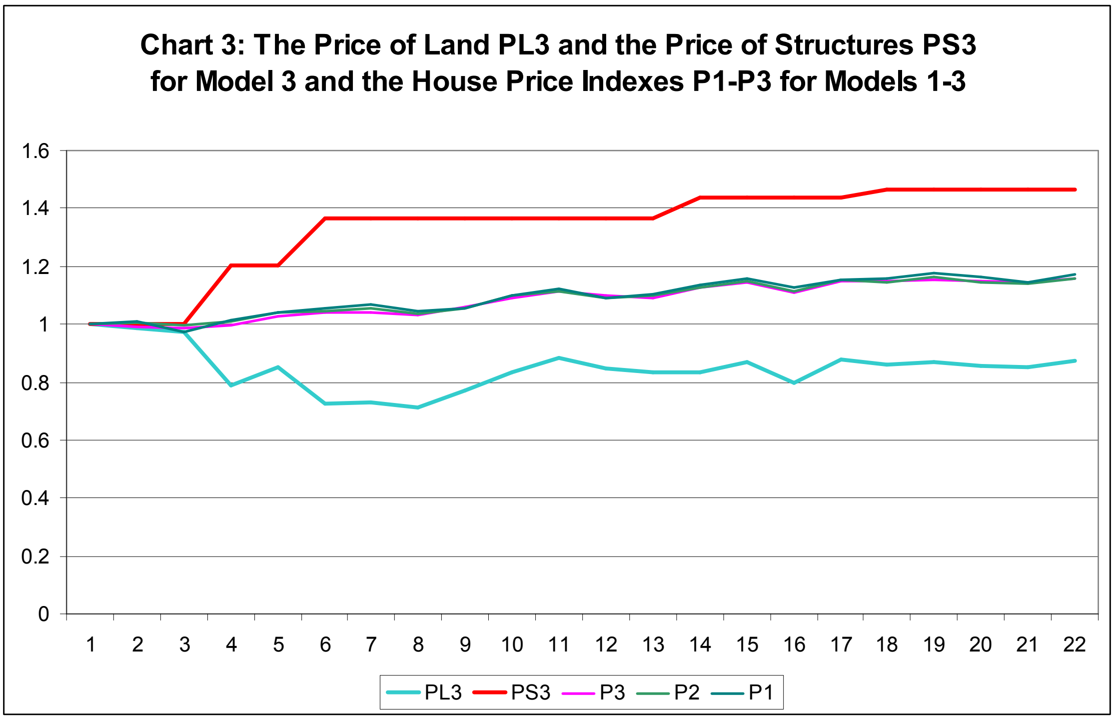
Table 5: The Price of Land P_{L3}, the Price of Structures P_{S3} and the Overall House Price Index P_3 Generated by Model 3 with the Corresponding Quantities Q_{L3}, Q_{S3} and Q_3

Quarter	P_{L3}	P_{S3}	P_3	Q_{L3}	Q_{S3}	Q_3
1	1.00000	1.00000	1.00000	11399.7	12021.1	23420.8
2	0.98653	1.00000	0.99337	11495.3	11509.0	23002.0
3	0.97504	1.00000	0.98743	12746.8	12025.8	24765.8
4	0.79044	1.20430	0.99424	13678.7	13731.3	27507.3
5	0.85400	1.20430	1.02581	12022.2	12108.5	24223.9
6	0.72682	1.36541	1.04072	14799.4	13781.0	28416.1
7	0.72965	1.36541	1.04216	10704.4	10838.2	21694.4
8	0.71084	1.36541	1.03311	12784.2	14054.8	27371.8
9	0.76921	1.36541	1.06024	12248.7	13654.2	26470.7
10	0.83364	1.36541	1.09053	12192.9	12974.6	25565.7
11	0.88184	1.36541	1.11368	12457.9	12961.2	25755.3
12	0.84946	1.36541	1.09782	13165.8	13195.3	26599.0
13	0.83508	1.36541	1.09067	13896.4	13816.6	27936.8
14	0.83274	1.43784	1.12574	14798.5	15583.4	30850.5
15	0.87166	1.43784	1.14447	12469.8	13011.2	25843.7
16	0.79730	1.43784	1.10822	12853.2	12966.1	26069.6
17	0.87833	1.43784	1.14766	12695.1	13293.4	26370.4
18	0.85918	1.46272	1.15034	18299.5	16853.7	35098.1
19	0.86830	1.46272	1.15504	11429.2	11191.2	22764.0
20	0.85825	1.46272	1.15010	15324.8	16055.2	31855.1
21	0.85001	1.46272	1.14613	12064.3	12563.1	24980.5
22	0.87545	1.46272	1.15823	14575.4	16079.0	31322.8
Mean	0.87545	1.3362	1.0844	13096	13376	26722

The imposition of monotonicity restrictions on the price of structures has led to smoother prices for land and structures but now, the price of structures P_{S3} grows too rapidly and the price of land P_{L3} falls too much.[34]

In the following Chart, we plot the new series P_{L3}, P_{S3} and P_3 along with the overall price house price indexes, P_1 and P_2, generated by Models 1 and 2 in the previous two sections.

[34] The mean price for P_{L3} was .87545 whereas the means for the unrestricted models P_{L1} and P_{L2} were 1.0941 and 0.94842 respectively. The mean price for P_{S3} was 1.3362 whereas the means for P_{S1} and P_{S2} were 1.0617 and 1.2310 respectively. Thus it seems clear that the imposition of monotonicity restrictions on the price of structures leads to structure prices which grow too rapidly and land prices which fall too much.



From Chart 3, it can be seen that the overall house price indexes generated by the three models considered thus far approximate each other very closely; i.e., the plots of P_1, P_2 and P_3 can hardly be distinguished from each other.[35] However, Chart 3 also indicates that Model 3 also fails to generate sensible land and structure price indexes: it is not credible that the price of a new structure could increase 36% over Quarters 3 to 6. Thus in the following section, we will use exogenous information on new construction prices in the Netherlands in an attempt to generate reasonable price series for land and structures.

5. Model 4: The Use of Exogenous Information on New Construction Prices

Many countries have national or regional new construction price indexes available from the national statistical agency on a quarterly basis.[36] This is the case for the Netherlands.[37] Thus if we are willing to 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, the statistical agency information on construction costs can be used to eliminate the multicollinearity problems that we encountered in the previous sections.

---

[35] The correlation coefficients between P1-P2, P1-P3 and P2-P3 were 0.9942, 0.9866 and 0.9947 respectively.

[36] 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 be used.

[37] From the Dutch Central Bureau of Statistics 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,...,p^{22}. As will be seen in Table 6 below, to our surprise, this series did not trend upwards monotonically.

Recall equations (5)-(7) in section 3 above. These equations are the estimating equations for the unrestricted regression model based on costs of production. In the present section, the constant quality house price parameters, the β^t for t = 2,...,22 in (5)-(7), are replaced by the following numbers, which involve only the single unknown parameter β^1:

$$(16)\ \beta^t = \beta^1 p^t \,; \qquad t = 2,3,...,22$$

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 unity in quarter 1. This new regression *Model 4* is again defined by equations (5)-(7) 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.[38] Thus the number of parameters to be estimated in this new restricted model is 68 as compared to the Model 2 number, which was 89.

Using the data for the town of "A", the estimated decade depreciation rate was $\delta^* = 0.1026$ (0.00448). The R^2 for this model was .8723, a drop from the previous Model 2 and 3 R-squares of .8756 and .8745. The log likelihood was −16239.7, a substantial decrease of 30.1 over the previous monotonicity restricted Model 3. The first period parameter values for the 3 marginal prices for land are $\alpha_S^{1*} = 0.1827$ (0.0256), $\alpha_M^{1*} = 0.3480$ (0.0640) and $\alpha_L^{1*} = 0.17064$ (0.0311). The first period parameter value for quality adjusted structures is $\beta^{1*} = 1.0735$ (0.0275) or 1073.5 Euros/m^2 which is substantially higher than the corresponding Model 1, 2 and 3 estimates which were 972.1, 882.9 and 815.9 Euros/m^2 respectively. Thus the imposition of a nationwide growth rate on the change in the price of quality adjusted structures for the town of "A" has had some effect on our previous estimates for the levels of land and structures prices.

As usual, we used equations (8)-(14) in order to construct a chained Fisher index of land prices, which we denote by P_{L4}. This index is plotted in Chart 4 and listed in Table 6 below. As was the case for the previous three models, the estimated period t price for a square meter of quality adjusted structures is β^{t*} (which in turn is now equal to $\beta^{1*}p^t$) 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 combined with the three land price and quantity series to form a chained *overall Fisher house price index* P_4 which is graphed in Chart 4 and listed in Table 6. The *constant quality structures price index* P_{S4} (a normalization of the series $\beta^{1*},...,\beta^{22*}$) is also found in Chart 4 and Table 6. It should be noted that the quarter to quarter movements in P_{S4} coincided with the quarter to quarter movements in the Statistics Netherlands New Dwellings Building Cost Price Index.

[38] This type of hedonic model that makes use of construction price information is similar to that introduced by Diewert (2010).

Table 6: The Price of Land P_{L4}, the Price of Structures P_{S4} and the Overall House Price Index P_4 Generated by Model 4 with the Corresponding Quantities Q_{L4}, Q_{S4} and Q_4

Quarter	P_{L4}	P_{S4}	P_4	Q_{L4}	Q_{S4}	Q_4
1	1.00000	1.00000	1.00000	7446.9	15749.3	23196.2
2	0.99248	1.01613	1.00842	7602.4	15073.6	22671.1
3	0.99248	1.00000	0.99769	8622.7	15752.1	24366.2
4	1.04399	0.99194	1.01035	9172.6	17988.4	27138.6
5	1.14791	0.98387	1.04007	8057.7	15868.4	23904.1
6	1.20958	0.95968	1.04554	9898.8	18057.6	28026.7
7	1.22438	0.96774	1.05593	7200.3	14201.1	21364.1
8	1.11160	1.00000	1.04056	8659.1	18424.2	26956.4
9	1.20134	0.98387	1.05818	8285.6	17899.5	26048.9
10	1.35900	0.97690	1.10428	8221.2	17006.0	25161.9
11	1.36491	0.99881	1.12097	8406.4	16988.4	25373.0
12	1.24923	1.02271	1.09813	8842.9	17298.3	26169.8
13	1.33155	0.99084	1.10504	9338.7	18106.5	27488.3
14	1.40580	1.00080	1.13646	9931.0	20429.3	30275.2
15	1.47191	0.99582	1.15492	8436.9	17050.8	25454.5
16	1.35274	0.99881	1.11711	8633.4	16994.4	25649.0
17	1.44763	1.01773	1.16136	8566.5	17421.0	25944.6
18	1.39479	1.02769	1.14980	12262.7	22082.7	34613.1
19	1.40183	1.05159	1.16770	7709.2	14659.1	22456.3
20	1.32049	1.07449	1.15549	10337.4	21044.1	31382.3
21	1.25610	1.09540	1.14825	8141.9	16465.8	24614.7
22	1.31144	1.09540	1.16627	9853.6	21082.3	30881.3
Mean	1.2541	1.0114	1.0928	8801.3	17529	26324

It can be seen that the price of structures does not behave in a monotonic manner but 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 variance of the land price series was much higher. The price of land peaked in Quarter 15, approximately 47% higher than the Quarter 1 level and then it generally trended downwards to finish 31% higher in Quarter 22. The results for this model look very reasonable since we expect the price of land to fluctuate much more than the price of structures.

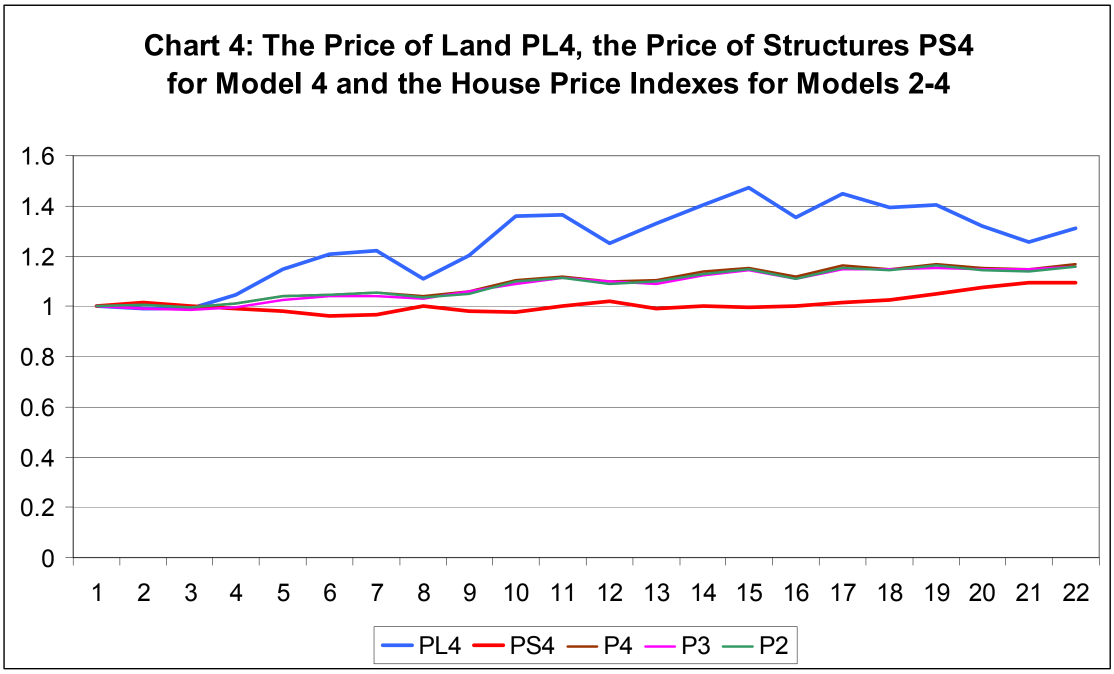


Chart 4 plots the price of land P_{L4} and structures P_{S4} for Model 4 along with the overall house price index generated by this model, P_4. We also plot the overall house price indexes generated by Models 2 and 3, P_2 and P_3, and compare these indexes with P_4. It can be seen that P_2, P_3 and P_4 can barely be distinguished as separate series in Chart 4.

Although the present model seems satisfactory, in the following section, we explore how the model can be improved by using additional information on housing characteristics.

6. 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 large structures have a higher quality than small structures; i.e., larger houses are likely to use more expensive construction materials than smaller houses. Thus it seems likely that using the same type of spline setup, but on S rather than L, we could improve the fit in our regression model. However, splining structures in the same way that we splined land would lead to a fairly complicated regression model and would add an additional 44 parameters. A much simpler alternative to using spline techniques on structures is to use 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 should increase as the number of rooms increases.[39] However, it should be noted that some housing experts believe that the price should decline as the structure size increases so the issue is not settled.[40]

Our regression *Model 5* is defined by equations (5)-(7) again except that the terms involving the quantity of structures, $\beta^t(1 - \delta A_n{}^t)S_n{}^t$ in each of the equations (5)-(7), are now replaced by the terms $\beta^1 p^t(1 - \delta A_n{}^t)(1 + \gamma R_n{}^t)S_n{}^t$ where β^1, δ and γ are parameters to be estimated, p^t is the Statistics Netherlands New Dwelling Construction Cost Price Index for quarter t described in the previous section, $A_n{}^t$ 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 that if γ equals 0, then the present model reduces to Model 4 in the previous section. Thus the present model has 69 parameters compared to the 68 parameters for Model 4. A priori, we expect the new parameter γ to be positive; i.e., as the number of rooms increases, we expect the price per m^2 of construction to also increase.

The R^2 for this model was .8736, an increase from the previous Model 4 R^2 of .8723. The log likelihood was −16222.6, a substantial increase of 17.1 over the previous Model 4 for the addition of only one new parameter, the room size parameter γ. The estimated decade depreciation rate was $\delta^* = 0.1089$ (0.00361). The first period parameter values for the 3 marginal prices for land are $\alpha_S{}^{1*} = 0.2207$ (0.0249), $\alpha_M{}^{1*} = 0.3465$ (0.0560) and $\alpha_L{}^{1*} = 0.1741$ (0.0307). The first period parameter value for quality adjusted structures is $\beta^{1*} = 1.0069$ (0.0212) or 1006.9 Euros/m^2. Note that this is the estimated construction cost for a new building (per meter squared) with four rooms in Quarter 1. Thus this new estimated Q1 building cost is not comparable to the Q1 building costs estimated by the previous model, since the earlier estimates applied to all houses irrespective of the number of rooms, which ranged from 2 to 14. The smallest t statistic was 4.64 for $\alpha_M{}^{3*}$ so that all parameters were significantly different from 0. The estimated number of rooms parameter was $\gamma^* = 0.02759$ (0.00493). Thus the estimated increase in the price of a new structure per m^2 in Quarter 1 due to an additional room is 0.02759/1.0069, which equals 2.74%. Thus the average premium in construction costs per m^2 in Quarter 1 of a 10 room house over a 2 room house is 2.74% times 8, which is 21.9% per m^2. This seems to be a reasonable quality premium.

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

[40] Palmquist (1984; 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; 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.

As usual, we used equations (8)-(14) in order to construct a chained Fisher index of land prices, which we denote by P_{L5}. This index is plotted in Chart 5 and listed in Table 7 below. The estimated quarter t price 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 $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \delta^* A_n^t)(1+\gamma^* R_n^t)S_n^t$.[41] The structures 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_5 which is graphed in Chart 5 and listed in Table 7. The *constant quality structures price index* P_{S5} (a normalization of the series $\beta^{1*},...,\beta^{22*}$) is also found in Chart 5 and Table 7.

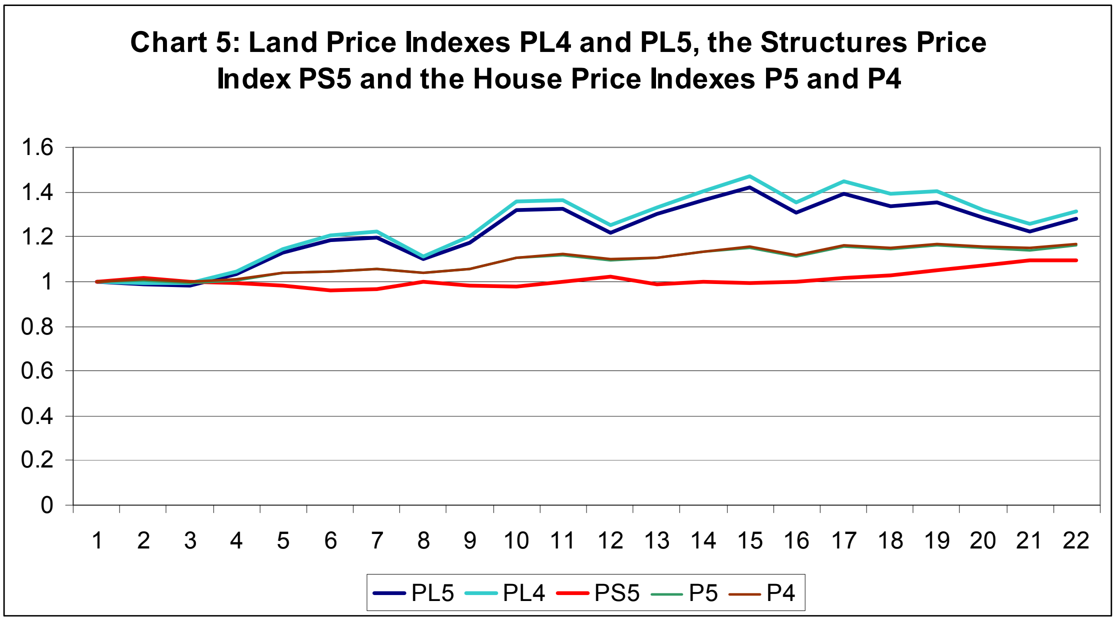
Table 7: The Price of Land P_{L5}, the Price of Structures P_{S5} and the Overall House Price Index P_5 Generated by Model 5 with the Corresponding Quantities Q_{L5}, Q_{S5} and Q_5

Quarter	P_{L5}	P_{S5}	P_5	Q_{L5}	Q_{S5}	Q_5
1	1.00000	1.00000	1.00000	8372.8	14816.2	23189.0
2	0.98919	1.01613	1.00626	8499.4	14218.0	22712.6
3	0.98251	1.00000	0.99362	9540.5	14929.4	24459.1
4	1.03180	0.99194	1.00760	10215.5	17005.8	27202.2
5	1.12890	0.98387	1.03902	8980.2	14954.4	23917.7
6	1.18484	0.95968	1.04555	10954.2	17004.3	28021.2
7	1.19793	0.96774	1.05555	8001.5	13397.9	21364.3
8	1.10152	1.00000	1.04067	9690.9	17363.8	26942.9
9	1.17454	0.98387	1.05632	9263.9	16952.2	26090.3
10	1.31868	0.97690	1.10370	9171.9	16053.3	25167.3
11	1.32326	0.99881	1.11928	9385.7	16035.6	25405.8
12	1.21947	1.02271	1.09563	9832.3	16368.2	26222.3
13	1.30263	0.99084	1.10718	10380.5	17003.3	27429.8
14	1.36153	1.00080	1.13530	11027.1	19376.0	30305.0
15	1.41932	0.99582	1.15332	9406.9	16109.7	25486.2
16	1.30854	0.99881	1.11409	9591.6	16114.3	25712.5
17	1.39053	1.01773	1.15633	9544.0	16562.4	26054.1
18	1.33811	1.02769	1.14266	13605.8	21006.0	34825.5
19	1.35373	1.05159	1.16328	8590.0	13876.1	22540.2

[41] 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 4 rooms. The appropriate price (per m^2) for a new house with 5, 6, ..., 14 rooms would be $\beta^{1*}p^t(1+\gamma^*)$, $\beta^{1*}p^t(1+2\gamma^*)$, ... , $\beta^{1*}p^t(1+10\gamma^*)$ and the price for a new house with 2 and 3 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. Thus the second method is equivalent to the first method.

20	1.28629	1.07449	1.15240	11516.1	19960.1	31464.8
21	1.22226	1.0954	1.14219	9075.5	15670.2	24739.9
22	1.28276	1.0954	1.16410	11009.9	19980.8	30933.9
Mean	1.2236	1.0114	1.0906	9802.6	16580	26372

It can be seen that the structures price series P_{S5} coincides with the structures price series P_{S4} for the previous model. This makes sense because both models impose the same rates of change on quality adjusted structures prices (equal to the Statistics Netherlands rates of change). Thus in Chart 5, we do not plot separately P_{S4} and P_{S5} since they are identical series.



From viewing Chart 5, it can be seen that our new model that allows for a quality adjustment for the construction of larger houses generates a somewhat different series for the price of land as compared to Model 4; i.e., P_{L5} lies below P_{L4} for Quarters 2-22. Note that P_{S5} is exactly equal to P_{S4} and the overall house price indexes, P_4 and P_5, are virtually identical[42]; i.e., they are difficult to distinguish in Chart 5.[43]

[42] The correlation coefficient between P_4 and P_5 is .99942.

[43] If P_4 almost equals P_5 and P_{S4} is exactly equal to P_{S5}, one might ask how can P_{L4} and P_{L5} differ so much? The answer is that while the rates of growth in the *price* of constant quality structures is the same in models 4 and 5, 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 4, the initial levels of land and constant quality structures was 7446.9 and 15749.3. Using Model 5, the initial levels of land and constant quality structures was 8372.8 and 14816.2. Thus going from Model 4 to 5, 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_4 and P_5 are virtually equal and the structures indexes P_{S4} and P_{S5} are exactly equal, it can be seen that these facts will imply that P_{L5} must grow more slowly than P_{L4}.

Recall that 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. However, running the regressions associated with Models 1-5, there were additional outliers (i.e., observations with large error terms), which were not deleted. This non deletion of regression outliers could affect our estimated coefficients, particularly if the outliers are either mostly positive or mostly negative. To determine whether outliers are a problem with Model 5, we looked at the empirical distribution of the resulting error terms for this model. We constructed 10 error intervals: $e_n^t < -100$[44]; $-100 \leq e_n^t < -75$; $-75 \leq e_n^t < -50$; ... ; $75 \leq e_n^t < 100$; $100 \leq e_n^t$. The number of observations that fell into these 10 bins was as follows: 9, 10, 57, 333, 1358, 1297, 319, 64, 34 and 6. Thus the empirical distribution of error terms appears to be fairly symmetric with a relatively small number of very large in magnitude errors.

Our conclusion at this point is that Model 5 is a satisfactory hedonic housing regression model that decomposes house prices into sensible land and structures components. The quality adjustments to the quantity of structures for the age of the structure and for the number of rooms also seem to be reasonable. The overall fit of the model also seems to be satisfactory: an R^2 of .8736 for such a small number of characteristics is quite good.[45]

The builder's model that we developed here could be further modified to take into account additional characteristics but a certain amount of careful thought is required so that the effects of introducing additional characteristics reflect the realities of housing construction and locational effects.[46] These construction realities will determine the appropriate functional form for the hedonic regression.

In the following two sections, we ignore the advice offered in the previous paragraph and we run more traditional *time dummy* linear and log linear regressions which enter the characteristics in a linear fashion, using the same data set that is used in this section. Thus we want to compare the performance of more traditional hedonic housing regression models with our builder's model: basically, we want to determine if all of the somewhat complicated functional forms we introduced for the builder's hedonic regression model are worth the bother. In section 7, we look at time dummy models using price as the dependent variable whereas in sections 8 and 9, we use the logarithm of price as the dependent variable.

7. Time Dummy Hedonic Regression Models using Price as the Dependent Variable

[44] Thus if an observation belonged to this bin, the associated error term was less than −100,000 Euros; recall that we measure house prices in thousands of Euros when running our regressions.

[45] However, the Dutch data may not be representative of other data sets where there could be more heterogeneity due to geography or differences in the types of houses being built over time.

[46] 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 lead to 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.

In this section, we will compare some "traditional" time dummy hedonic regression models for housing where the selling price is the dependent variable using our data set. In the following two sections, we will use the log of the selling price as our dependent variable.

As indicated in section 2 above, there are reasons to believe that the selling price of a property is linearly related to the plot area of the property plus the area of the structure due to the competitive nature of the house building industry. If the age of the structure is treated as another characteristic that has an importance in determining the price of the property, then the following *linear time dummy hedonic regression model* might be an appropriate one:

(17) $V_n{}^t = \alpha + \beta L_n{}^t + \gamma S_n{}^t + \delta A_n{}^t + \tau^t + \varepsilon_n{}^t$; $t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$

Note the differences between (2) and (17): (2) did not have a constant term, the age variable A interacted with structures variable S in a nonlinear fashion in (2) and, more importantly, in (2), the coefficients associated with land and structures were time dependent whereas in (17), these parameters are constant over time. On the other hand, (17) has introduced 21 new parameters τ^t (τ^1 is set equal to 0), which shift the hedonic surface in a parallel fashion over each quarter.

The above linear regression model was run using our data for the town of "A". The R^2 for this *Model 6* was .8539 and the log likelihood was −16473.1. Holding characteristics constant and neglecting error terms, the *difference in price* for a house with the same characteristics turns out to be constant across any two periods, but the relative price for the same model will not in general be constant. Thus an overall index will be constructed which uses the prices generated by the estimated parameters in (17) and evaluated at the sample average amounts of L, S and the average age of a house A. The resulting quarterly house prices for this "average" model were converted into an index, P_6, which is listed in Table 8 below and plotted in Chart 6.

The hedonic regression model defined by (17) neglects the fact that the interaction of age with the selling price of the property likely takes place via a multiplicative interaction with the structures variable and not via a general additive factor. Thus we rerun the present model but using quality adjusted structures as an explanatory variable rather than just entering age A as a separate stand alone characteristic. Thus (17) is replaced with the following *nonlinear time dummy hedonic regression model*:

(18) $V_n{}^t = \alpha + \beta L_n{}^t + \gamma(1-\delta A_n{}^t)S_n{}^t + \tau^t + \varepsilon_n{}^t$; $t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$

The R^2 for this *Model 7* was .8625 and the log likelihood was −16366.9.[47] Thus treating the age variable A as a quality adjustment variable for the quantity of structures S has led to an increase in R^2 over the previous traditional linear model and an increase in log

[47] The estimates for the key parameters were as follows: $\alpha^* = -9.4034$ (2.3099), $\beta^* = 0.26305$ (0.004066), $\gamma^* = 1.0674$ (0.01859) and $\delta^* = 0.10885$ (0.003603). These estimates are similar to the corresponding estimates we obtained for Model 2. Note that this model can be run as a linear regression model.

likelihood of 106.2. Thus a more sensible model in the terms of a priori theory has led to a big increase in log likelihood. As in the previous model, holding characteristics constant and neglecting error terms, the difference in price for a house with the same characteristics is constant across any two periods, but the relative price for the same model will not in general be constant. Thus an overall index is constructed using the prices generated by the estimated parameters in (18) and evaluated at the sample average amounts of L, S and the average age of a house A. The resulting quarterly house prices for this "average" model were converted into an index, P_7, which is listed in Table 8 below and plotted in Chart 6.

We now introduce the number of rooms, R, as an additional explanatory variable for the simple linear regression model that was defined by (17). The new model is the following *linear time dummy hedonic regression model* with age A and the number of rooms R as linear explanatory variables[48]:

(19) $V_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \phi R_n^t + \tau^t + \varepsilon_n^t$; $\quad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$

The R^2 for this *Model 8* was .8548 and the log likelihood was −16462.1, an increase of 11.0 from the Model 6 log likelihood. Thus the addition of the new room parameter ϕ is statistically significant and improves the fit of the initial model defined by (17). As was the case for the previous two models, holding characteristics constant and neglecting error terms, the difference in price for a house with the same characteristics is constant across any two periods, but not the ratio of prices. Thus an overall index will be constructed which uses the prices generated by the estimated parameters in (19) and evaluated at the sample average amounts of L, S and the average age of a house A. The resulting quarterly house prices for this "average" model were converted into an index, P_8, which is listed in Table 8 below and plotted in Chart 6.

We now treat the R and A variables as quality adjustment variables for S. Thus our new *nonlinear time dummy hedonic regression model* with age A and the number of rooms R as nonlinear quality adjustment factors is the following regression model:

(20) $V_n^t = \alpha + \beta L_n^t + \gamma(1-\delta A_n^t)(1+\phi R_n^t)S_n^t + \tau^t + \varepsilon_n^t$; $\quad t = 1,...,22$; $n = 1,...,N(t);\ \tau^1 \equiv 0.$

The R^2 for this *Model 9* was .8651 and the log likelihood was −16333.8, a jump of 128.3 over the log likelihood for its strictly linear counterpart in Model 8.[49] Again, we interpret this very large jump in log likelihood for the present model over its traditional linear counterpart as support for treating the A and S variables in a nonlinear fashion in

---

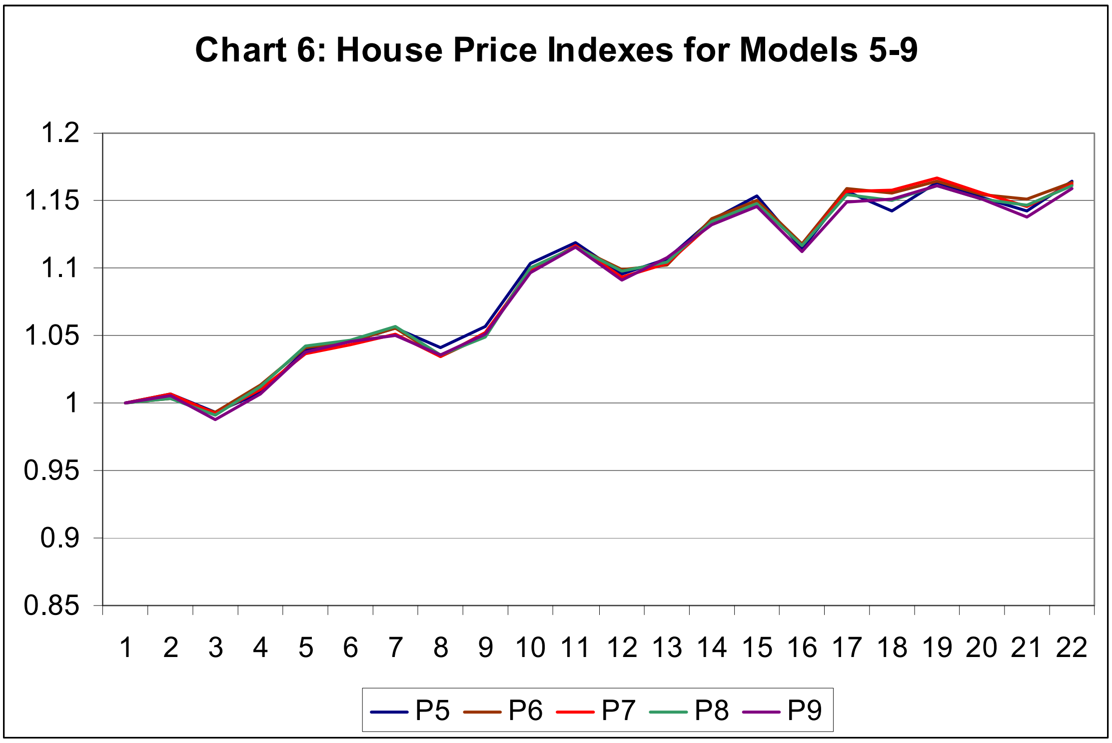
[48] The variable R is equal to the original number of rooms variable less 4 in this regression and the following ones. Thus R equals 0 for houses with 4 rooms.

[49] The estimates for the key parameters were as follows: $\alpha^* = 0.8563$ (1.7459), $\beta^* = 0.26351$ (0.003910), $\gamma^* = 0.96027$ (0.0200), $\delta^* = 0.12034$ (0.003449) and $\phi^* = 0.04329$ (0.005425). Thus Model 9 generates a relatively high depreciation rate of 12.03% per decade and a relatively high extra room premium on the constant quality structures price equal to 0.04329/0.96027 = 4.51% per extra room. We do not regard these estimates as being as accurate as the corresponding estimates obtained using Model 5 (because Model 5 allows the price of land and structures to vary independently during each quarter while the present time dummy model does not allow for independent movements in the prices of land and structures over time).

accordance with our a priori economic reasoning. The log likelihood for Model 9 is also a jump of 33.1 over the log likelihood in the nonlinear Model 7, which is the same as Model 9 except that Model 9 has added the room variable R as a quality adjustment variable. Thus the extra room parameter ϕ is definitely significantly different from zero. As in the previous models in this section, an overall house price index is constructed using the prices generated by generating a price series for an average model using the estimated parameters in (20) evaluated at the sample average amounts of L, S, the average age of a house A and the average number of rooms (less 4). The resulting quarterly house prices for this "average" model were converted into an index, P_9, which is listed in Table 8 below and plotted in Chart 6.

Table 8: Overall House Price Indexes Generated by Models 5-9

Quarter	P_5	P_6	P_7	P_8	P_9
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00626	1.00388	1.00694	1.00325	1.00591
3	0.99362	0.99371	0.99201	0.99066	0.98745
4	1.00760	1.01280	1.01007	1.01169	1.00678
5	1.03902	1.04151	1.03710	1.04238	1.03749
6	1.04555	1.04570	1.04347	1.04644	1.04505
7	1.05555	1.05610	1.05130	1.05638	1.05016
8	1.04067	1.03463	1.03459	1.03506	1.03538
9	1.05632	1.04964	1.05273	1.04855	1.05098
10	1.10370	1.10027	1.09813	1.09962	1.09718
11	1.11928	1.11606	1.11670	1.11563	1.11611
12	1.09563	1.09851	1.09366	1.09812	1.09130
13	1.10718	1.10276	1.10292	1.10499	1.10757
14	1.13503	1.13645	1.13368	1.13487	1.13195
15	1.15332	1.15015	1.14823	1.14799	1.14600
16	1.11409	1.11821	1.11649	1.11614	1.11245
17	1.15633	1.15916	1.15665	1.15482	1.14916
18	1.14266	1.15508	1.15744	1.15043	1.15164
19	1.16328	1.16485	1.16629	1.16156	1.16159
20	1.15240	1.15483	1.15535	1.15120	1.15157
21	1.14219	1.15056	1.14599	1.14632	1.13825
22	1.16410	1.16311	1.16231	1.16085	1.15914
Mean	1.0906	1.0913	1.0901	1.0899	1.0879



From the above Chart, it can be seen that all five indexes are fairly close to each other. Thus the conclusion that can be drawn from this section is that fairly simple time dummy linear regression models using our 3 or 4 main characteristics (L, S, A and R) of the houses sold in the Dutch town of "A" can capture the same trends in overall house prices as our more complex builder's Model 5.

Hedonic housing regressions using the logarithm of price as the dependent variable are much more popular in the housing literature than models that use price as the dependent variable. Thus in the following section, we look at time dummy hedonic regression models that use the logarithm of price as the dependent variable.

8. Time Dummy Hedonic Regression Models using the Logarithm of Price as the Dependent Variable and Linear Independent Variables

The most popular hedonic regression models regress the log of the price of the good on *either* a *linear function* of the characteristics or of *the logs of the characteristics* along with time dummy variables.[50] We will consider the log-linear models in this section and log-log models in the following section.

The four models that we will consider in this section are entirely analogous to Models 6-9 in the previous section except now, the natural logarithm of the selling price, $\ln V_n^t$, replaces V_n^t in the regression models defined by equations (17)-(20). This replacement leads to Models 10-13. Thus *Model 10* is defined as follows:

[50] This methodology was developed by Court (1939; 109-111) as his hedonic suggestion number two.

(21) $\ln V_n{}^t = \alpha + \beta L_n{}^t + \gamma S_n{}^t + \delta A_n{}^t + \tau^t + \varepsilon_n{}^t$; $t = 1,...,22$; $n = 1,...,N(t)$; $\tau^1 \equiv 0$.

Using our usual data set, the R^2 for this model was .8308 and the log likelihood was 1977.7. Note that this R^2 and log likelihood are not comparable to the R^2 and log likelihoods for our previous model, since the dependent variables have changed. In order to obtain measures of fit that are comparable to our earlier R^2 measures, we generated predicted values for our dependent variables using the estimated regression coefficient and exponentiated these predicted values, thus obtaining predicted values for the $V_n{}^t$. We then calculated the correlation coefficient between these predicted values for the $V_n{}^t$ and the actual $V_n{}^t$. The square of this correlation coefficient is our measure of fit that is comparable to our earlier measures of fit. We denote this *levels measure of fit* as R^{*2}. The R^{*2} for Model 10 turned out to be .7884, much less that the R^2 for the comparable regression that used price as the dependent variable, Model 6 (where the R^2 was .8539).

Note that if we exponentiate both sides of (21) and neglect the error term, then the house price $V_n{}^t$ would equal $e^\alpha\,[\exp L_n{}^t]^\beta\,[\exp S_n{}^t]^\gamma\,[\exp A_n{}^t]^\delta\,[\exp\tau^t]$. Thus if we could observe a house with the *same characteristics* in two consecutive periods t and t+1, the corresponding price relative (neglecting error terms) would equal $[\exp\tau^{t+1}]/[\exp\tau^t]$ and this can serve as the chain link in a price index. Thus it is particularly easy to construct a house price index using this model; see Table 9 and Chart 7 below for the resulting index which is labelled as P_{10}. The fact that all houses with the same characteristics generate the same price index is one of the main advantages of models that use the logarithm of price as the dependent variable.[51]

Model 11 is defined by equations (18) except that $\ln V_n{}^t$, replaces $V_n{}^t$ as the dependent variable. Thus this model is defined by the following nonlinear regression model:

(22) $\ln V_n{}^t = \alpha + \beta L_n{}^t + \gamma(1-\delta A_n{}^t)S_n{}^t + \tau^t + \varepsilon_n{}^t$; $t = 1,...,22$; $n = 1,...,N(t)$; $\tau^1 \equiv 0$.

Thus this model is similar to the previous Model 10 except now the age variable A acts as a nonlinear quality adjustment variable for the structures quantity variable S instead of just appearing in the regression in a linear fashion.

Using our usual data set, the R^2 for this model was .8266 and the levels R^{*2} was .7884. The log likelihood for Model 11 was 1935.2, a drop of 42.5 from the log likelihood for Model 10. Thus it appears that the addition of more structure on the functional form for the hedonic regression led to a model which fit less well. However, both Models 10 and 11 have the same problem: the underlying value of a house is assumed to be proportional to a *multiplicative function* of the plot size L and the area of the structure S (instead of being proportional to an *additive function* of L and S or quality adjusted S).[52] Not only

[51] The other major advantage of a log price specification is that the resulting stochastic specification is more reasonable; i.e., exponentiating both sides of (21) shows that the exponential of the original error term is proportional to the predicted price so that houses with big prices will tend to have big error terms, which is very reasonable.

[52] Francke (2008; 168) bluntly noted that this multiplicative specification made no sense.

are Models 10 and 11 dubious from the viewpoint of the builder's approach to modeling house prices, these two models fit the data much less well than *all* of the other models considered thus far. An overall house price series is formed by exponentiating the time dummy variables τ^{t*} as was done for Model 10 and the resulting series is denoted by P_{11}; see Table 9 and Chart 7 below for the resulting index.

Model 12 adds the room variable R[53] as a linear explanatory variable to Model 10. Thus this model is defined by equations (23) below:

$$(23)\ \ln V_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \phi R_n^t + \tau^t + \varepsilon_n^t\ ; \qquad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

The R^2 for this model was .8337 and the levels R^{*2} was .7934. The log likelihood for Model 12 was 2008.0, an increase of 30.3 over the log likelihood for Model 10. Thus the addition of the extra room parameter ϕ is statistically very significant. However, the fit for this model is not good as compared to Models 1-9. An overall house price series was formed by exponentiating the time dummy variables τ^{t*} as was done for Model 10 and the resulting series is denoted by P_{12}; see Table 9 and Chart 7 below for the resulting index.

Model 13 adds the number of rooms variable R as a nonlinear quality adjustment variable to Model 11. Thus this model is defined by equations (24) below:

$$(24)\ \ln V_n^t = \alpha + \beta L_n^t + \gamma(1-\delta A_n^t)(1+\phi R_n^t)S_n^t + \tau^t + \varepsilon_n^t\ ; \quad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

The R^2 for this model was .8271 and the levels R^{*2} was .7827. The log likelihood for Model 13 was 1940.0, an decrease of 68.0 over the log likelihood for Model 12 but an increase of 4.8 of the log likelihood for Model 11, which is the same as the present model with ϕ set equal to 0. Thus the addition of the extra room parameter ϕ is statistically significant. Again, the fit for this model is not good as compared to Models 1-9. An overall house price series was formed by exponentiating the time dummy variables τ^{t*} as was done for Model 10 and the resulting series is denoted by P_{13}; see Table 9 and Chart 7 below for the resulting index.[54]

Table 9: Overall House Price Indexes Generated by Models 5 and 10-13

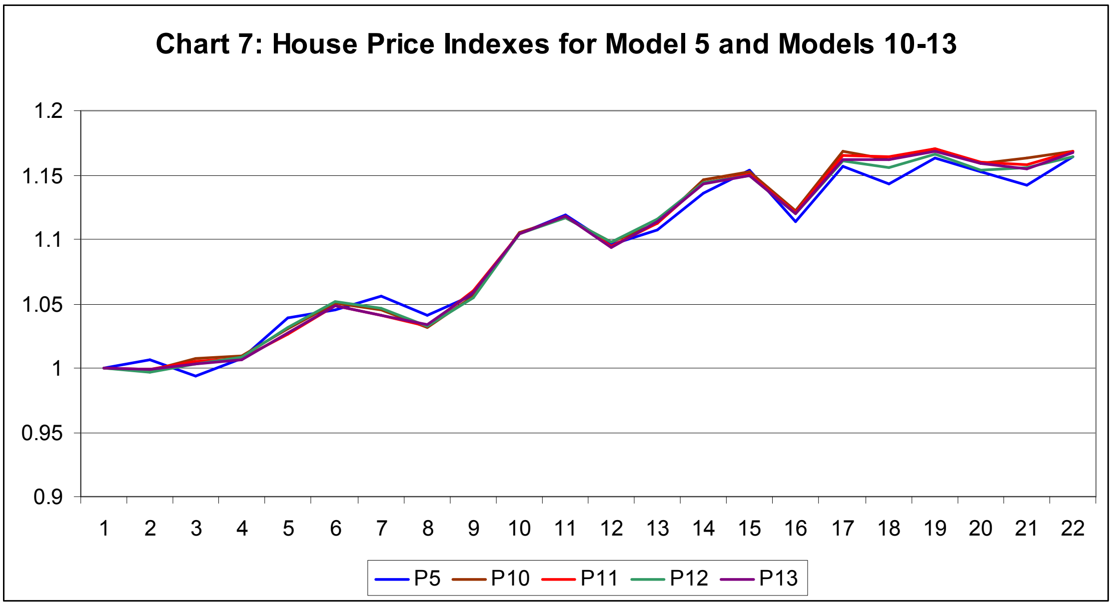
Quarter	P_5	P_{10}	P_{11}	P_{12}	P_{13}
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00626	0.99760	0.99913	0.99674	0.99869
3	0.99362	1.00744	1.00496	1.00314	1.00342
4	1.00760	1.00998	1.00692	1.00843	1.00582
5	1.03902	1.03035	1.02672	1.03168	1.02689
6	1.04555	1.05035	1.04836	1.05154	1.04893

[53] As usual, R_n^t is equal to the number of rooms in house n sold during quarter t less 4. Thus R_n^t takes on integer values between −2 and 10.

[54] We do not report any parameter estimates for Models 10-13 since we do not believe that these models are very realistic. Nevertheless, these Models do generate overall house price indexes which are reasonably close to our preferred index P_5.

7	1.05555	1.04571	1.04095	1.04623	1.04058
8	1.04067	1.03162	1.03299	1.03231	1.03332
9	1.05632	1.05597	1.05992	1.05447	1.05930
10	1.10370	1.10546	1.10399	1.10465	1.10370
11	1.11928	1.11695	1.11832	1.11652	1.11816
12	1.09563	1.09785	1.09436	1.09746	1.09365
13	1.10718	1.11217	1.11235	1.11587	1.11404
14	1.13530	1.14684	1.14423	1.14460	1.14353
15	1.15332	1.15283	1.15067	1.14967	1.14986
16	1.11409	1.12248	1.12107	1.11948	1.11963
17	1.15633	1.16812	1.16514	1.16140	1.16232
18	1.14266	1.16259	1.16393	1.15539	1.16167
19	1.16328	1.17084	1.17069	1.16581	1.16886
20	1.15240	1.15939	1.16007	1.15384	1.15863
21	1.14219	1.16286	1.15799	1.15631	1.15510
22	1.16410	1.16794	1.16853	1.16461	1.16740
Mean	1.0906	1.0943	1.0932	1.0923	1.0924

It can be seen that the series P_{10}-P_{13} lie above our preferred builder's model overall index P_5 on average but as Chart 7 shows, at times this relationship is reversed. The bottom line is that simple log linear models, in spite of their a priori functional form implausibility, can approximate our preferred index reasonably closely but the degree of approximation is not quite as close as was the case for the simple linear models discussed in the previous section. It can also be seen that the addition of the room variable does not change the overall indexes very much.



It can be seen that at times, the deviation of the log linear models P_{10}-P_{13} from P_5 can be noticeable.

9. Time Dummy Hedonic Regression Models using the Logarithm of Price as the Dependent Variable and the Logs of Structures and Land as Independent Variables

We will now consider counterparts to the models in the previous section, except that instead of entering the independent variables S and L in a linear fashion, we will enter them as the *logarithms* of S and L.

Thus *Model 14* is defined as follows:

$$(25)\ \ln V_n^t = \alpha + \beta \ln L_n^t + \gamma \ln S_n^t + \delta A_n^t + \tau^t + \varepsilon_n^t\ ; \qquad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

The R^2 for this model was .8552 and the levels R^{*2} was .8769, which is the highest R^{*2} thus far (recall that the level R^{*2} for the corresponding log-linear Model 10 was only .7884). The log likelihood for Model 14 was 2249.2, an increase of 30.3 over the log likelihood for Model 10. Thus the switch to the log-log model from the log-linear model dramatically improves the fit.[55]

Note that if we exponentiate both sides of (25) and neglect the error term, then the house price V_n^t would equal $e^\alpha\,[L_n^t]^\beta\,[S_n^t]^\gamma\,[\exp A_n^t]^\delta\,[\exp\tau^t]$. Hence if we could observe a house with the *same characteristics* in two consecutive periods t and t+1, then the corresponding price relative (neglecting error terms) would equal $[\exp\tau^{t+1}]/[\exp\tau^t]$ and this can serve as the chain link in a price index. Thus again, it is particularly easy to construct a house price index using this log-log model; see Table 10 and Chart 8 below for the resulting index which is labelled as P_{14}. Note also that if L_n^t and S_n^t increase by a scalar factor $\lambda > 0$, then V_n^t would equal $e^\alpha\,[\lambda L_n^t]^\beta\,[\lambda S_n^t]^\gamma\,[\exp A_n^t]^\delta\,[\exp\tau^t]$, which is $\lambda^{\beta+\gamma}$ times the initial value of V_n^t. Thus if $\beta + \gamma = 1$, then house prices exhibit constant returns to scale in the land and structure characteristics and decreasing returns if $\beta + \gamma < 1$. Our empirical results indicate decreasing returns to scale in these characteristics.

Model 15 is an alternative version of the previous model where we enter quality adjusted structures in log form. Thus this model is defined by the following nonlinear regression model:

$$(26)\ \ln V_n^t = \alpha + \beta \ln L_n^t + \gamma \ln[(1-\delta A_n^t)S_n^t] + \tau^t + \varepsilon_n^t\ ; \qquad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

This model is similar to the previous Model 14 except now, the age variable A acts as a nonlinear quality adjustment variable for the structures quantity variable S instead of just appearing in the regression in a linear fashion.

The R^2 for this model was .8532 and the levels R^{*2} was .8758, a bit lower than the previous model measures of fit. The log likelihood for Model 15 was 2224.8, a drop of

[55] The key parameter estimates were as follows: $\alpha^* = 0.41513$ (0.05292), $\beta^* = 0.41008$ (0.006110), $\gamma^* = 0.52886$ (0.01334) and $\delta^* = -0.07386$ (0.001867).

24.4 from the log likelihood for Model 14.[56] Thus it appears that the addition of more economic structure on the functional form for the hedonic regression led to a model which fit less well. However, both Models 14 and 15 have the same problem: the underlying value of a house is assumed to be proportional to a *multiplicative function* of the plot size L and the area of the structure S (instead of being proportional to an *additive function* of L and S or quality adjusted S). Thus these models are not consistent with the builder's approach to modeling housing hedonics. As was the case with all previous time dummy models that used the log of price as the dependent variable in the regression, an overall house price series can be formed by exponentiating the time dummy variables τ^{t*} and the resulting series is denoted by P_{15}; see Table 10 and Chart 8 below for the resulting index.

Although Model 15 cannot be justified from our builder's perspective, it is possible to give a justification for the model from the viewpoint of a *household or purchaser oriented approach* to valuing the services of housing. We will now outline this consumer perspective approach, which is essentially due to Muth (1971) and McMillen (2003).[57]

We postulate that purchasers of houses have the same (cardinal) utility function, $f(z_1,z_2)$, that aggregates the amounts of two relevant characteristics, $z_1 > 0$ and $z_2 > 0$, into the overall utility of the housing model with characteristics z_1, z_2 into the scalar welfare measure, $f(z_1,z_2)$. Thus purchasers will prefer model 1 with characteristics z_1^1,z_2^1 to model 2 with characteristics z_1^2,z_2^2 if and only if $f(z_1^1,z_2^1) > f(z_1^2,z_2^2)$.[58] Generally speaking, having more of every characteristic is always preferred by purchasers. The next assumption that we make is that in period t, there is a positive generic price for all models (once they have been quality adjusted for the amounts of characteristics that they possess), ρ^t, such that the household's *willingness to pay*, $W^t(z_1,z_2)$, for a model with characteristics z_1 and z_2 is equal to the generic model price ρ^t times the utility generated by the model, $f(z_1,z_2)$; i.e., we have for each model n with characteristics z_{1n}^t, z_{2n}^t that is purchased in period t, the following *willingness to pay* for model n:

(27) $W^t(z_{1n}^t,z_{2n}^t) = \rho^t f(z_{1n}^t,z_{2n}^t)$.

In order to relate the above model to sales in the Dutch city of "A", identify the first characteristic with the size of the land area of the house n sold in period t, L_n^t, and the second characteristic with the quality adjusted (for the age of the structure) size of the structure, $S_n^{t*} \equiv (1 - \delta A_n^t)S_n^t$, where S_n^t is the *unadjusted size of the structure*, δ is the *depreciation rate* for structures and A_n^t is the *age of the structure*. Finally, set the

---

56 The estimates for the key parameters were as follows: $\alpha^* = 0.33975$ (0.05238), $\beta^* = 0.40969$ (0.006382), $\gamma^* = 0.54114$ (0.01311) and $\delta^* = 0.10187$ (0.003089). Since $\beta^* + \gamma^* = .951 < 1$, we have decreasing returns to scale in land and quality adjusted structures.

57 For more elaborate justifications for household based hedonic regression models, see Muellbauer (1974) and Diewert (2003). Muth (1971) originally set up his model as a production function model and other contributors to this literature like Rosen (1978) and Thorsnes (1998) followed his example but McDonald (1981) showed that the same model could be reinterpreted in a consumer context.

58 It is natural to impose some regularity conditions on the characteristics aggregator function f like continuity, monotonicity (if each component of the vector z^1 is strictly greater than the corresponding component of z^2, then $f(z^1) > f(z^2)$ and $f(0,0) = 0$.

willingness to pay for the housing unit, $W^t(L_n^t,(1-\delta A_n^t)S_n^t)$, equal to the *selling price of the property*, V_n^t and we have the following hedonic regression model:

(28) $V_n^t = \rho^t f(L_n^t,(1-\delta A_n^t)S_n^t)$.

There remain the problems of choosing a stochastic specification for the hedonic regression model (28) and of choosing a functional form for the hedonic utility function f. The simplest choices for $f(L,S^*)$ are that (i) f is a *linear function* of L and S^* or (ii) f has a *Cobb-Douglas* functional form. These two choices lead to the following hedonic regression models after adding independently distributed normal errors ε_n^t with means zero and constant variances to each choice of functional form:[59]

(29) $V_n^t = \rho^t(\alpha L_n^t + \beta(1-\delta A_n^t)S_n^t) + \varepsilon_n^t$; $n = 1,...,N(t); t = 1,...,T$;
(30) $\ln V_n^t = \ln\rho^t + \alpha + \beta \ln L_n^t + \gamma \ln[(1-\delta A_n^t)S_n^t] + \varepsilon_n^t$; $n = 1,...,N(t); t = 1,...,T$.

In order to identify all of the parameters, we require a normalization on the hedonic prices ρ^t. It is natural to set ρ^t equal to one in the first period:

(31) $\rho^1 = 1$.

It can be seen that the hedonic regression model defined by (29) and (31) is essentially a reparameterization of our first simple regression model explained in section 2 above (with some important additional restrictions on the parameters). Finally, note that the Cobb-Douglas model defined by (30) and (31) is essentially a reparameterization of the model defined by equations (26) with τ^t equal to $\ln\rho^t$. Thus McMillan has provided a purchaser's justification for Model 15.[60]

Note that the builder's models explained in sections 2-6 above were able to generate separate estimates for the price of land and for the price of quality adjusted structures whereas the present model does not seem to be able to generate these separate estimates. However, it is possible to use the log-log model (or any other hedonic model based on a hedonic utility function $f(z_1,z_2)$) in order to generate *imputed estimates* for the price of

[59] Note that the linear $f(L,S^*)$ that is defined in (29) is linearly homogeneous in the variables L and S^*. The Cobb-Douglas f that is defined in (30) will be linearly homogeneous in the characteristics if $\beta + \gamma = 1$. McMillen's (2003) hedonic housing model that uses a consumer perspective is essentially the Cobb-Douglas model defined by (30).

[60] However, Model 15 can also be given a supply side justification using the cost of production approach due to Muth (1971) and Thorsnes (1997; 101). These authors assumed that instead of equation (28), the value of a property under consideration in period t, V^t, is equal to the price of housing output in period t, ρ^t, times the quantity of housing output H(L,K) where the *production function* H is a CES function. Thus Thorsnes assumed that $V^t = \rho^t H(L,K) = \rho^t [\alpha L^\sigma + \beta K^\sigma]^{1/\sigma}$ where ρ^t, σ, α and β are parameters , L is the lot size of the property and K is the amount of structures capital in constant quality units (the counterpart to our S^*). As σ approaches 0, this model will approach the Cobb-Douglas model defined by (30) with a suitable relabeling of the parameters. Our problem with this model is that there is only one independent time parameter ρ^t in it for each period t, whereas our builder's model has two, β^t and γ^t, which allows the price of land and structures to vary freely between periods. Essentially, our builder's model assumes that contractors can purchase lots and build houses in an independent manner.

land, $\rho_L{}^t$, and for quality adjusted structures, $\rho_S{}^t$. The basic idea is to take the consumer's period t willingness to pay function, $W^t(z_1,z_2)$, and differentiate it with respect to z_1 and z_2. These two partial derivatives will give us estimates of the consumer's increase in well being in period t, valued at the period t price for the hedonic aggregate, due to a marginal increase in the quantities of z_1 and z_2; i.e., this procedure generates imputed prices for extra units of z_1 and z_2 in period t.[61] Thus we define $\rho_L{}^t$ and $\rho_S{}^t$ as follows:

$$(32)\ \rho_L{}^t \equiv \partial W^t(z_1{}^{t*},z_2{}^{t*})/\partial z_1 = \rho^t \partial f(z_1{}^{t*},z_2{}^{t*})/\partial z_1\ ; \qquad t = 1,...,22;$$
$$(33)\ \rho_S{}^t \equiv \partial W^t(z_1{}^{t*},z_2{}^{t*})/\partial z_2 = \rho^t \partial f(z_1{}^{t*},z_2{}^{t*})/\partial z_2\ ; \qquad t = 1,...,22$$

where $z_1{}^{t*}$ and $z_2{}^{t*}$ are the average amounts of land and quality adjusted structures for the properties sold in period t.[62] We use (32) and (33) to generate imputed price series for land and quality adjusted structures, using our estimated coefficients for α, β, γ and δ in order to form an estimated $f(z_1,z_2)$ function. The land structures price series defined by (32) and (33) were normalized to equal 1 in quarter 1 and are listed as P_{L15} and P_{S15} in Table 11 and they are plotted in Chart 9 below.

Model 16 adds the room variable R[63] as a linear explanatory variable to Model 14. Thus this model is defined by equations (34) below:

$$(34)\ \ln V_n{}^t = \alpha + \beta \ln L_n{}^t + \gamma \ln S_n{}^t + \delta A_n{}^t + \phi R_n{}^t + \tau^t + \varepsilon_n{}^t\ ; \quad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

The R^2 for this model was .8562 and the levels R^{*2} was .8784, which is the best fit of all the models considered thus far. The log likelihood for Model 16 was 2261.3, an increase of 12.1 over the log likelihood for Model 14, which did not include the R variable.[64] An overall house price series was formed by exponentiating the time dummy variables τ^{t*} as was done for Model 14 and the resulting series is denoted by P_{16}; see Table 10 and Chart 8 below for the resulting index.

Model 17 adds the number of rooms variable R as a nonlinear quality adjustment variable to Model 15. Thus this nonlinear model is defined by equations (35) below:

$$(35)\ \ln V_n{}^t = \alpha + \beta \ln L_n{}^t + \gamma \ln[(1-\delta A_n{}^t)(1+\phi R_n{}^t)S_n{}^t] + \tau^t + \varepsilon_n{}^t\ ; \qquad t = 1,...,22;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

The R^2 for this model was .8538 and the levels R^{*2} was .8769. The log likelihood for Model 17 was 2232.8, an decrease of 28.5 over the log likelihood for Model 16 but an

---

[61] This methodology was initially developed in Diewert, Haan and Hendricks (2010).

[62] After determining the quarterly average amounts of land and quality adjusted structures in our sample, the overall mean of these average amounts was 257.76 m^2 for land (min = 235.05, max = 293.77) and 103.04 for quality adjusted structures (min = 98.63, max = 108.90). Thus the quarterly average amounts of land varied much more than the quarterly average amounts of quality adjusted structures.

[63] As usual, $R_n{}^t$ is equal to the number of rooms in house n sold during quarter t less 4. Thus $R_n{}^t$ takes on integer values between –2 and 10.

[64] The key parameter estimates were as follows: $\alpha^* = 0.55193$ (0.05963), $\beta^* = 0.41006$ (0.006090), $\gamma^* = 0.49907$ (0.01461), $\delta^* = -0.075168$ (0.001879) and $\phi^* = 0.014142$ (0.002877).

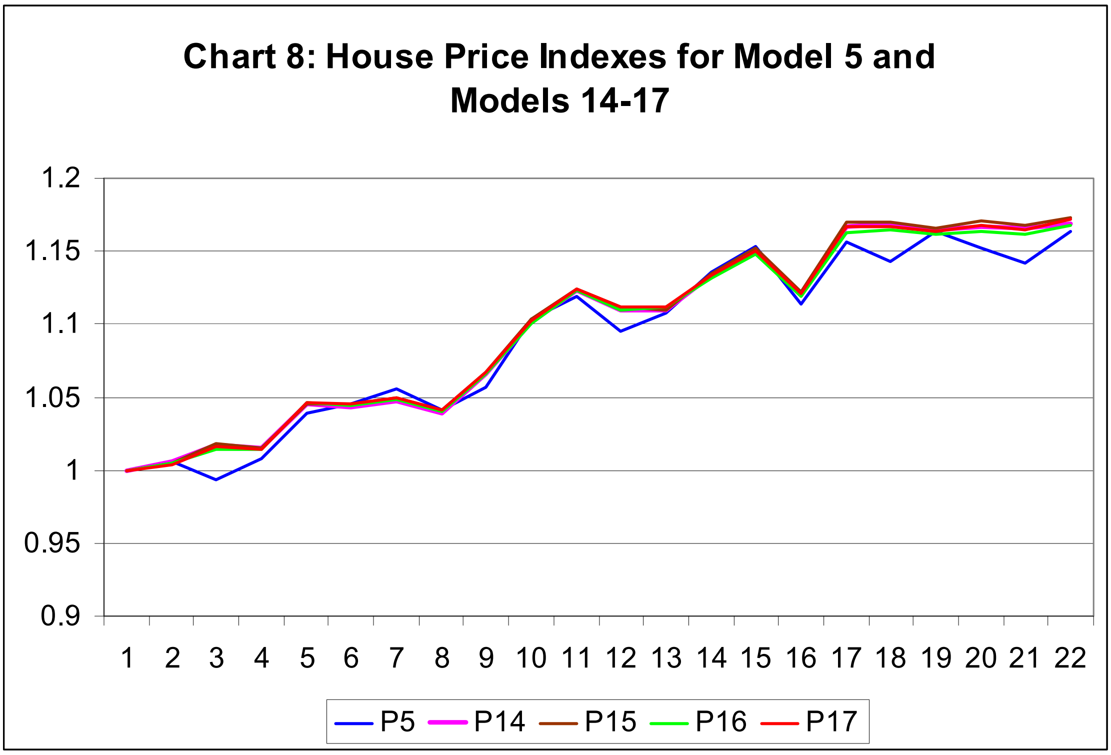
increase of 8.0 over the log likelihood for Model 15, which is the same as the present model with ϕ set equal to 0. Thus the addition of the extra room parameter ϕ is statistically significant.[65] An overall house price series was formed by exponentiating the time dummy variables τ^{t*} as was done for Model 15 and the resulting series is denoted by P_{17}; see Table 10 and Chart 8 below for the resulting index.

Define the *quality adjusted amount of structures* for house n in quarter t as $S_n^{t*} \equiv (1-\delta^* A_n^t)(1+\phi^* R_n^t)S_n^t$. With this new definition for quality adjusted structures, the methodology surrounding equations (32) and (33) above can be repeated for our new model that has added the number of rooms as a quality adjustment variable for the quantity of structures. Again use (32) and (33) with the new definition of z_2^t to generate imputed price series for land and quality adjusted structures, using our estimated coefficients for α, β, γ, δ and ϕ in order to form an estimated $f(z_1,z_2)$ function. The land structures price series defined by (32) and (33) were normalized to equal 1 in quarter 1 and are listed as P_{L17} and P_{S17} in Table 11 and they are plotted in Chart 9 below.

Table 10: Overall House Price Indexes Generated by Models 5 and 14-17

Quarter	P_5	P_{10}	P_{11}	P_{12}	P_{13}
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00626	1.00590	1.00474	1.00531	1.00416
3	0.99362	1.01698	1.01812	1.01450	1.01614
4	1.00760	1.01490	1.01507	1.01400	1.01433
5	1.03902	1.04546	1.04564	1.04603	1.04613
6	1.04555	1.04344	1.04489	1.04422	1.04562
7	1.05555	1.04779	1.04910	1.04801	1.04939
8	1.04067	1.03942	1.04103	1.03983	1.04148
9	1.05632	1.06658	1.06738	1.06558	1.06663
10	1.10370	1.10097	1.10302	1.10052	1.10283
11	1.11928	1.12297	1.12397	1.12291	1.12400
12	1.09563	1.10981	1.11145	1.10934	1.11125
13	1.10718	1.10929	1.10992	1.11161	1.11186
14	1.13530	1.13319	1.13476	1.13198	1.13385
15	1.15332	1.14941	1.15177	1.14767	1.15052
16	1.11409	1.12053	1.12206	1.11891	1.12087
17	1.15633	1.16704	1.17018	1.16307	1.16721
18	1.14266	1.16843	1.17025	1.16439	1.16708
19	1.16328	1.16458	1.16550	1.16182	1.16331
20	1.15240	1.16663	1.17047	1.16344	1.16815
21	1.14219	1.16525	1.16800	1.16138	1.16507
22	1.16410	1.16911	1.17314	1.16732	1.17197
Mean	1.0906	1.0967	1.0982	1.0955	1.0974

[65] The key parameter estimates were as follows: $\alpha^* = 0.43951$ (0.05795), $\beta^* = 0.40931$ (0.006185), $\gamma^* = 0.51941$ (0.01421), $\delta^* = 0.10584$ (0.003369) and $\phi^* = 0.022273$ (0.005869).



It can be seen that the overall price series generated by the various log-log models, P_{14}-P_{14}, approximate each other rather closely. However, they tend to be above our best overall index P_5 with some occasional noticeable differences. But, for the most part, the differences between P_5 and the indexes P_{14}-P_{17} are not large. This indicates that the log-log simple linear regressions can generate quite acceptable overall house price indexes, at least for our data set for the town of "A".[66]

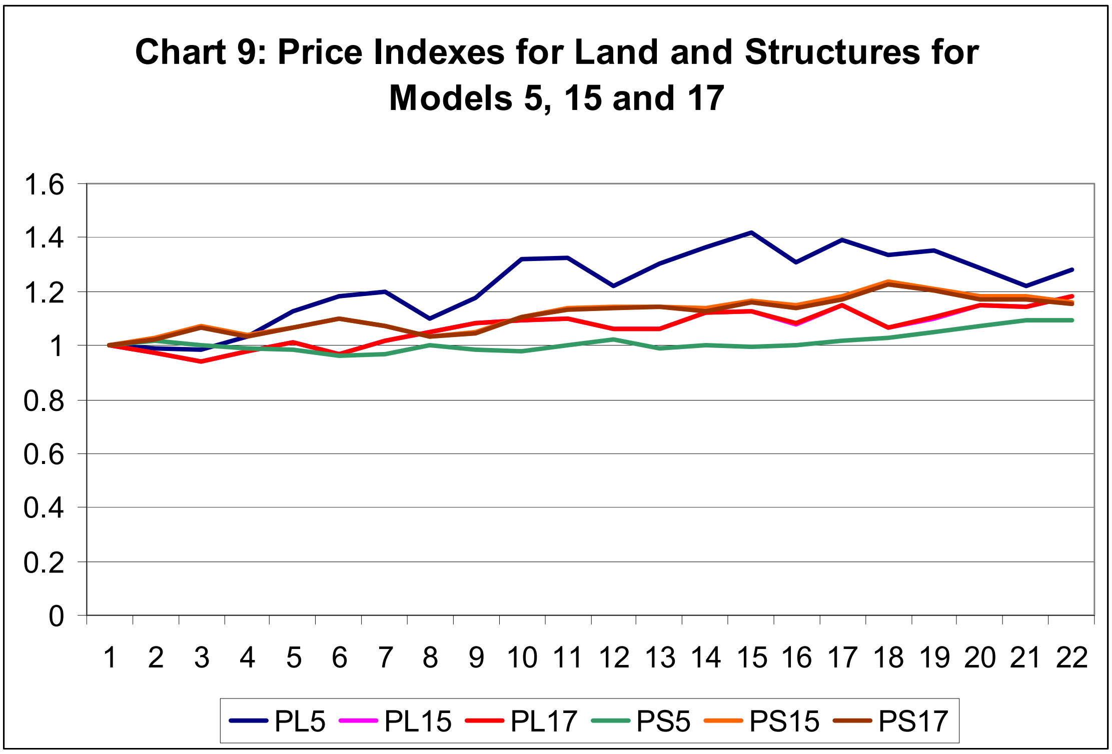
The land and structures price indexes generated by Models 15 and 17 are listed below in Table 11 and plotted in Chart 9. Chart 9 also includes our best builder model price indexes for land and structures generated by Model 5, which included the number of rooms variable.

Table 11: Prices for Land and Structures Generated by Models 5, 15 and 17

Quarter	P_{L5}	P_{L15}	P_{L17}	P_{S5}	P_{S15}	P_{S17}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.98919	0.97188	0.97210	1.01613	1.02660	1.02440
3	0.98251	0.94164	0.94207	1.00000	1.07173	1.06563
4	1.03180	0.97922	0.98011	0.99194	1.03908	1.03588
5	1.12890	1.00987	1.01018	0.98387	1.06813	1.06717
6	1.18484	0.96723	0.96734	0.95968	1.09869	1.09804
7	1.19793	1.01510	1.01596	0.96774	1.07111	1.06965

[66] The correlation coefficients of P_5 with P_{14}-P_{17} are as follows: 0.98839, 0.98813, 0.99028 and 0.98970 respectively.

8	1.10152	1.04744	1.04748	1.00000	1.03294	1.03190
9	1.17454	1.08311	1.08337	0.98387	1.04999	1.04499
10	1.31868	1.09360	1.09390	0.97690	1.10619	1.10377
11	1.32326	1.09877	1.09912	0.99881	1.13776	1.13539
12	1.21947	1.06299	1.06386	1.02271	1.14121	1.13730
13	1.30263	1.05940	1.05916	0.99084	1.14167	1.14328
14	1.36153	1.11891	1.11913	1.00080	1.13648	1.12973
15	1.41932	1.12696	1.12697	0.99582	1.16669	1.16286
16	1.30854	1.08035	1.08127	0.99881	1.14659	1.14058
17	1.39053	1.14937	1.15049	1.01773	1.18132	1.17267
18	1.33811	1.06786	1.06810	1.02769	1.23936	1.22863
19	1.35373	1.10216	1.10269	1.05159	1.21035	1.20473
20	1.28629	1.14897	1.14887	1.07449	1.17948	1.17192
21	1.22226	1.14233	1.14352	1.09540	1.18112	1.17171
22	1.28276	1.18427	1.18477	1.09540	1.15927	1.15346
Mean	1.2236	1.0660	1.0664	1.0114	1.1175	1.1133



It can be seen that the price of land series generated by Model 15 and 17, P_{L15} and P_{L17}, cannot be distinguished from each other on Chart 9 and the price of structures series, P_{S15} and P_{S17}, generated by these two models can barely be distinguished from each other. But the important point to note is that these four series approximate each other to a reasonable degree; i.e., all four series are for the most part *below* the price of land series P_{L5} and *above* the price of structures series P_{S5} generated by our best builder's model. Thus *these* *time dummy models tend to generate price series for land and quality adjusted structures that show the same general trends.* This will almost always be the case for time dummy hedonic regression models. We do not think that this pattern of price movements is as realistic as the corresponding price series generated by our best builder's models, where the price of land tends to fluctuate much more than the price of structures.

A problem with the hedonic regression models discussed in this paper is that they are not immediately suitable for use by statistical agencies that have to produce real time indexes that cannot be revised. Thus as the data for a new quarter are added to an existing data set, a new hedonic regression of the type discussed in this paper could be run, leading to changing historical index values. A simple solution to this difficulty is available. First, one chooses a "suitable" number of periods (equal to or greater than two) where it is thought that the hedonic regression model will yield "reasonable" results; this will be the *window length* (say M periods) for the sequence of regression models that will be estimated. Second, an initial regression model is estimated and the appropriate indexes are calculated using data pertaining to the first M periods in the data set. Next, a second regression model is estimated where the data consist of the initial data less the data for period 1 but adding the data for period M+1. Appropriate price indexes are calculated for this new regression model but only the rate of increase of the index going from period M to M+1 is used to update the previous sequence of M index values. This procedure is continued with each successive regression dropping the data of the previous earliest period and adding the data for the next period, with one new update factor being added with each regression. If the window length is a year, then this procedure is called a *rolling year hedonic regression model* and for a general window length, it is called a *rolling window hedonic regression model*. This is exactly the procedure used recently by Shimizu, Takatsuji, Ono and Nishimura (2010) and Shimizu, Nishimura and Watanabe (2011) in their hedonic regression models for Tokyo house prices.[67] Diewert (2010) tested out this method using some of the data for the town of "A" and found that it worked well in the sense that the rolling window approach generated indexes that were very similar to those generated by a single regression over the entire sample period.

10. Conclusion

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

- 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, then a wide variety of hedonic regression models that use these variables seem to generate much the same *overall* house price indexes, except that the log linear model fits the data more poorly.
- 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

[67] An analogous procedure has also been used by Ivancic, Diewert and Fox (2011) and Haan and van der Grient (2011) in their adaptation of the GEKS method for making international comparisons in the scanner data context.

statistical agency information on the price movements of new dwelling units, generated satisfactory results for our data set.

- Time dummy hedonic regression models can be used to generate land and structures price series (as was shown in section 9) but the resulting estimates are not satisfactory since the time dummy model artificially forces the movements in the prices of land and structures to be similar when we know that the price of land is much more volatile than the price of structures.
- Adding the number of rooms in the dwelling unit as an explanatory variable in our hedonic regressions did improve the fit but did not change the indexes substantially.
- Splining land also improved the fit of our hedonic regressions and led to somewhat smoother land price indexes in our best builder's model.
- It is important to delete observations in the regressions which are range outliers.

Some topics for follow up research include the following:

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

References

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

Clapp, J.M. (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.

Court, A.T. (1939), "Hedonic Price Indexes with Automotive Examples", pp. 99-117 in *The Dynamics of Automobile Demand*, New York: General Motors Corporation.

Crone, T.M., L.I. Nakamura and R.P. Voith (2009), "Hedonic Estimates of the Cost of Housing Services: Rental and Owner Occupied Units", pp. 67-84 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

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.

Davis, M.A. and M.G. Palumbo (2008), "The Price of Residential Land in Large US Cities", *Journal of Urban Economics* 63, 352-384.

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

Diewert, W.E. (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, V6T 1Z1.

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., J. de Haan and R. Hendriks (2010), "The Decomposition of a House Price Index into Land and Structures: A Hedonic Regression Approach", Discussion Paper 10-01, Department of Economics, University of British Columbia, Vancouver, Canada, V6T1Z1.

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

Diewert, W.E., A.O. Nakamura and L.I. Nakamura (2009), "The Housing Bubble and a New Approach to Accounting for Housing in A CPI", *Journal of Housing Economics* 18, 156-171.

Fleming, M.C. and J.G. Nellis (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", 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.

Glaeser, E.L. and J. Gyourko (2003), "The Impact of Building Restrictions on Housing Affordability", *Economic Policy Review* 9, 21-39.

Gouriéroux, C. and A. Laferrère (2009), "Managing Hedonic House Price Indexes: The French Experience", *Journal of Housing Economics* 18, 206-213.

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.

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

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

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

Hill, R.J. (2011), "Hedonic Price Indexes for Housing", Statistics Directorate, Working Paper No 36, February 14, Paris: OECD.

Hill, R.J., D. Melser and I. Syed (2009), "Measuring a Boom and Bust: The Sydney Housing Market 2001-2006", *Journal of Housing Economics* 18, 193-205.

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

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

McDonald, J.F. (1981), "Capital-Land Substitution in Urban Housing: A Survey of Empirical Estimates", *Journal of Urban Economics* 9, 190-211.

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.

Muellbauer, J. (1974), "Household Production Theory, Quality and the 'Hedonic Technique'", *American Economic Review* 64, 977-994.

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* 66:3, 394-404.

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.

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* 16:3, 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., H. Takatsuji, H. Ono and Nishimura (2010), “Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices”, *International Journal of Housing Markets and Analysis* 3:4, 351-368.

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

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