Decomposing an RPPI into Land and Structures Components

8

Introduction

8.1 In Chapter 3 it was mentioned that for national accounts and CPI purposes, it will be useful or necessary to have a decomposition of the residential property price index (RPPI) into two components: a quality adjusted price index for structures and a price index for the land on which the house is built. The present chapter outlines how hedonic regression can be utilized to derive such a decomposition. Hedonic regression methods were discussed in Chapter 5.

8.2 Some economic reasoning will be helpful to derive an appropriate hedonic regression model. Think of a property developer who is planning to build a structure on a particular property. He or she will likely determine the selling price of the property after the structure is completed by first calculating the total expected cost. This cost will be equal to the floor space area of the structure, say S square meters, times the building cost per square meter, γ say, plus the cost of the land, which will be equal to the cost per square meter, β say, times the area of the land site, L. We follow a *cost of production approach* to modeling the property price. That is, the functional form for the hedonic price function is assumed to be determined by the supply side of the market, i.e., by independent contractors.[1]

8.3 Now consider a sample of properties of the same general type, which have structure areas S_n^t and land areas L_n^t in period t for $n = 1,...,N(t)$; the prices p_n^t are equal to costs of the above types plus error terms ε_n^t which are assumed to have means 0. This gives rise to the following hedonic regression model for period t where β^t and γ^t are the parameters to be estimated:[2]

$$p_n^t = \beta^t L_n^t + \gamma^t S_n^t + \varepsilon_n^t \qquad (8.1)$$

$$t = 1,...,T \; ; \; n = 1,...,N(t)$$

The quantity of land L_n^t and the quantity of structures S_n^t associated with the sale of property n in period t are the only two property characteristics included in this very simple model; the corresponding prices in period t are the price of a square meter of land β^t and the price of a square meter of structure floor space γ^t. Separate linear regressions of the form (8.1) can be performed for each time period t in the sample.

8.4 The "builder's model" (8.1) essentially relates to newly-built dwellings. To make it applicable to existing

or resold houses we should account for the fact that older structures will be worth less than newer structures due to *depreciation* of the structures. Information on the age of the structure will therefore be needed. The next section shows how depreciation can be incorporated into the model, similar to what was done in the examples for the town of "A" presented in Chapter 5. It will also be shown how additional land and structures characteristics can be included as explanatory variables.

Accounting for Depreciation and Additional Characteristics

Depreciation

8.5 Suppose that in addition to information on the selling price of property n at time period t, p_n^t, the land area of the property, L_n^t, and the structure area, S_n^t, information on the age of the structure at time t, say A_n^t, is available. If straight line depreciation is assumed, the following model is a straightforward extension of (8.1) to include "existing" houses:

$$p_n^t = \beta^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t \qquad (8.2)$$

$$t = 1,...,T \; ; \; n = 1,...,N(t)$$

where the parameter δ reflects the (straight line) *depreciation rate* as the structure ages one additional period. If structure age is measured in years, δ will probably be between 0.5 % and 2 %. This will be an underestimate of "true" depreciation because it will not account for major renovations or additions to the structure. The estimated straight line depreciation rate in (8.2) should therefore be interpreted as a *net depreciation rate*; i.e., a gross depreciation rate less the rate of renovations and additions to the structure. Model (8.2) will not work for very old structures since, if they are still in use, they will likely have been extensively renovated.[3]

8.6 Notice that (8.2) is a *nonlinear* regression model whereas (8.1) is a linear regression model.[4] Because the depreciation parameter δ is regarded as fixed over time, (8.2) would have to be estimated as one nonlinear regression over all time periods in the sample, whereas model (8.1) can be run as a period by period linear regression. The period t price of land in model (8.2) will be the estimate for the parameter β^t and the price of a unit of a newly built structure for period t will be the estimate for

γ^t. The period t quantity of land for property n is L_n^t and the period t quantity of structures for property n, expressed in equivalent units of a new structure, is $(1-\delta A_n^t)S_n^t$, where S_n^t is the floor space area of property n in period t.

8.7 Expensive properties probably have relatively large absolute errors compared to inexpensive properties, so it might be better to assume multiplicative rather than additive errors. However, we prefer an additive model specification as the purpose is to decompose the aggregate value of housing into the *sum* of structures and land components; the use of additive errors facilitates this decomposition. When there is evidence of heteroskedasticity, weighted regressions can be considered. Several researchers suggested hedonic regression models that lead to additive decompositions of a property price into land and structures components. [5]

8.8 There is a potential problem with the above builder's model, namely multicollinearity. Large structures are generally built on large plots of land, so that S_n^t and L_n^t could be highly collinear (i.e., the land-structure ratios L_n^t / S_n^t could be centered around a constant). This could give rise to unstable estimates of the quality adjusted prices β^t and γ^t for land and structures. As will be seen in the example using data for the Dutch town of "A", the problems of multicollinearity and instability do indeed occur. In general, multicollinearity is not a major problem if the goal is to produce an overall house price index, but it is problematic if the goal is to produce separate price indices for land and structures components. Some possible methods for overcoming the multicollinearity problem will be suggested in later on.

8.9 The hedonic regression model (8.2) has the implication that the parameters would have to be re-estimated whenever the data for a new period became available. To overcome this problem, a "rolling window" approach could be applied. A suitable window length T would be chosen, [6] the model defined by (8.2) or (8.3) would be estimated using the data for the last T periods, and the existing series for price of land and for price of structures would be updated using the chain link factors β^T / β^{T-1} and γ^T / γ^{T-1}. This approach will be illustrated below.

Adding More Characteristics

8.10 The above basic nonlinear hedonic regression framework can be generalized to encompass the traditional array of characteristics used in real estate hedonic regressions. Suppose that we can associate with each property n transacted in period t a list of K characteristics

$X_{n1}^t, X_{n2}^t, ..., X_{nK}^t$ that are price determining characteristics for the land on which the structure was built and a similar list of M characteristics $Y_{n1}^t, Y_{n2}^t, ..., Y_{nM}^t$ that are price determining characteristics for the type of structure. The following equations generalize (8.2) to the present setup: [7]

$$p_n^t = \beta^t \left[1 + \sum_{k=1}^{K} X_{nk}^t \eta_k \right] L_n^t + \gamma^t (1 - \delta A_n^t) \left[1 + \sum_{m=1}^{M} Y_{nm}^t \lambda_m \right] S_n^t + \varepsilon_n^t$$
$$t = 1,...,T \; ; \; n = 1,...,N(t) \quad (8.3)$$

where the parameters to be estimated are now the K quality of land parameters, $\eta_1,...,\eta_K$, the M quality of structures parameters, $\lambda_1,...,\lambda_M$, the period t quality adjusted price for land β^t and the period t quality adjusted price for structures γ^t. The quality adjusted amount of land, L_n^{t*}, and the corresponding quality adjusted amount of structures, S_n^{t*}, for property n in period t are defined as follows:

$$L_n^{t*} \equiv \left[1 + \sum_{k=1}^{K} X_{nk}^t \eta_k \right] L_n^t \quad (8.4)$$

$$S_n^{t*} \equiv \left[1 + \sum_{m=1}^{M} Y_{nm}^t \lambda_m \right] S_n^t$$

$$t = 1,...,T \; ; \; n = 1,...,N(t)$$

8.11 To illustrate how X and Y variables can be formed, consider the list of explanatory variables in the hedonic housing regression model reported by Li, Prud'homme and Yu (2006; 23). The following variables in their list of explanatory variables can be viewed as variables that affect structures quality; i.e., they are Y type variables: number of bedrooms, number of bathrooms, number of garages, number of fireplaces, age of the unit, age squared of the unit, exterior finish is brick or not, dummy variable for new units, unit has hardwood floors or not, heating fuel is natural gas or not, unit has a patio or not, unit has a central built in vacuum cleaning system or not, unit has an indoor or outdoor swimming pool or not, unit has a hot tub unit or not, unit has a sauna or not, and unit has air conditioning or not. The following variables can be assumed to affect the quality of the land; i.e., they are X type location variables: unit is at the intersection of two streets or not (corner lot or not), unit is at a cul-de-sac or not, shopping center is nearby or not, and various suburb location dummy variables.

8.12 Equations (8.3) and (8.4) show how the quality adjusted amounts of land and structures would be calculated if the goal is to construct price indices for the *sales* of properties of the type that are included in the hedonic regression model. If the goal is to construct price indices for the *stock* of properties of the type included in the regression, then the construction of appropriate weights becomes more complex. These weighting problems will be discussed in the next section.

---

[5] See Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic, Longhofer and Redfearn (2007), Diewert (2007), Francke (2008), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert, de Haan and Hendriks (2010) (2011) and Diewert (2010).

[6] The model becomes a modified adjacent period hedonic regression model for T = 2.

[7] This generalization was suggested by Diewert (2007).

Aggregation and Weighting Issues: Indices for Sales versus Stocks of Housing

8.13 As was explained in Chapter 5, the construction of an RPPI for the sales of property using standard hedonic regression techniques is fairly straightforward. Typically, a separate hedonic regression of the type defined by (8.3) will be run for each locality or region in a country.[8] Recall that once a particular regression has been run, period t quality adjusted prices for land, P_L^t, and for structures, P_S^t, for the region under consideration can be defined in terms of the estimated parameters for the model as follows:

$$P_L^t \equiv \beta^t \qquad (8.5)$$
$$t = 1,...,T$$

$$P_S^t \equiv \gamma^t \qquad (8.6)$$
$$t = 1,...,T$$

The corresponding quality adjusted quantities of land and structures for the region, say Q_L^t and Q_S^t can also be defined in terms of the estimated parameters using definitions (8.4) above as follows:

$$Q_L^t \equiv \sum_{n=1}^{N(t)} L_n^{t*} = \sum_{n=1}^{N(t)} \left[1 + \sum_{k=1}^{K} X_{nk}^t \eta_k \right] L_n^t \qquad (8.7)$$
$$t = 1,...,T$$

$$Q_S^t \equiv \sum_{n=1}^{N(t)} S_n^{t*} = \sum_{n=1}^{N(t)} \left[1 + \sum_{m=1}^{M} Y_{nm}^t \lambda_m \right] S_n^t \qquad (8.8)$$
$$t = 1,...,T$$

8.14 If hedonic regressions, for say R regions, of the type defined by (8.3) have been run for the T periods of data, then the algebra associated with (8.5)-(8.8) can be repeated for each region r. Denote the resulting prices and quantities for region r that are the counterparts to (8.5)-(8.8) by P_{Lr}^t, P_{Sr}^t, Q_{Lr}^t and Q_{Sr}^t for $r = 1,...,R$ and $t = 1,...,T$. Now Fisher (sales) RPPIs for land can be constructed using the regional price and quantity data for land, $P_L^t \equiv [P_{L1}^t,...,P_{LR}^t]$ and $Q_L^t \equiv [Q_{L1}^t,...,Q_{LR}^t]$, for each time period t $(t = 1,...,T)$. Similarly, Fisher (sales) RPPIs for structures can be constructed using the price and quantity data for structures in each period t, $P_S^t \equiv [P_{S1}^t,...,P_{SR}^t]$ and $Q_S^t \equiv [Q_{S1}^t,...,Q_{SR}^t]$, for $t = 1,...,T$.[9]

8.15 As was the case with stratification methods, it is now necessary to consider how to construct an RPPI for the *stock* of residential properties when hedonic regression methods are used. The period t hedonic cell prices P_{Lr}^t and P_{Sr}^t defined by the region r counterparts to (8.5) and (8.6) can still be used as cell prices to construct stock price indices for land and structures, but the counterpart quantities Q_{Lr}^t and Q_{Sr}^t defined by (8.7) and (8.8) are no longer appropriate; these quantities need to be replaced by estimates that apply to the *total stock of dwelling units* in the region (or some other reference population) for regression r at time t, say Q_{Lr}^{t*} and Q_{Sr}^{t*}, for $r = 1,...,R$. Thus, the counterpart summations in (8.7) and (8.8) are now taken over the entire stock of dwellings in region r in period t instead of just the dwelling units that were sold in period t. Period t information on the quantity of land L_{nr}^t for every unit n in the region that is in scope for the hedonic regression model m is now required, along with the accompanying characteristics information X_{nrk}^t for every land characteristic k, as well as data on the quantity of the structures S_{nr}^t, along with the accompanying characteristics information Y_{nrm}^t for every structures characteristic m. With these new population quantity weights, the rest of the details of the index construction are the same as was the case for the sales RPPI.

8.16 In order to construct appropriate period t population stock weights, it will be necessary for the country to have census information on the housing stock with enough details on each dwelling unit in the stock so that the required information on the quantity of land and structures and the accompanying characteristics can be calculated. If information on new house construction (plus the required characteristics data) and on demolitions is available in a timely manner, the census information can be updated and period t estimates for the constant quality amounts of land and structures, the Q_{Lr}^{t*} and Q_{Sr}^{t*}, can be approximated in a timely manner. Hence, stock RPPIs for land and structures can be constructed using Fisher indices, as was the case for the sales RPPI. If timely data on new construction and demolitions is unavailable, it may only be possible to construct fixed base Laspeyres type price indices using the quantity weights from the last available housing census.

8.17 If census information is not available at all (or if data on the characteristics of the dwelling units is missing), it still may be possible to approximate RPPIs for land and structures using hedonic regression techniques. If characteristics data on the residential properties that are sold in each period is stored over a large period of time, an approximate distribution of dwelling units by type can be constructed. This information may then be used to approximate a stock based RPPI in the manner explained above.

---

[8] Separate hedonic regressions may also be run for different types of property as well as for different locations. However, cost considerations may mean that a comprehensive system of regressions covering all properties in the country cannot be implemented so that there will only be a sample of representative hedonic regressions. The aggregation issues in the sampling case are too complex to be considered here; the exact details for constructing a national index would depend on the nature of the sampling design.

[9] As was the case for stratification methods, fixed base or chained indices could be constructed. Rolling window hedonic regressions could also be run. The rolling window approach will be explained later.

Main Advantages and Disadvantages

8.18 This section summarizes the main advantages and disadvantages of using hedonic regression methods to construct an RPPI for land and structure components. The main advantages are:

- If the list of available property characteristics is sufficiently detailed, the method adjusts for both sample mix changes and quality changes of the individual houses.

- Price indices can be constructed for different types of dwellings and locations through a proper stratification of the sample. Stratification has a number of additional advantages.

- The method is probably the most efficient method for making use of the available data.

- The method is virtually the only method that can be used to decompose the overall price index into land and structures components.

8.19 The main disadvantages of the hedonic regression approach are:

- The method is data intensive since it requires data on all relevant property characteristics (in particular, the age, the type and the location of the properties in the sample as well as information on the structure and lot size) so it is relatively expensive to implement.

- The method may not lead to reasonable results due to multicollinearity problems.

- While the method is essentially reproducible, different choices can be made regarding the set of characteristics entered into the regression, the functional form for the model, the stochastic specification, possible transformations of the dependent variable, etc., which could lead to varying estimates of overall price change.

- The general idea of the hedonic method is easily understood but some of the technicalities may not be easy to explain to users.

Application on Data for the Town of "A": Preliminary Approaches

8.20 The general techniques explained in this chapter will now be illustrated using the data set for the Dutch town of "A", which was described at the end of Chapter 4.

We have data on sales of detached dwellings for 14 quarters, starting in the first quarter of 2005. Recall the notation used above and in Chapters 4 and 5: there were $N(t)$ sales of detached houses in quarter t, where p_n^t is the selling price of house n. There is information available on three characteristics: area of the plot in square meters, L_n^t; floor space area of the structure in square meters, S_n^t; and age in decades of house n in period t, A_n^t.

The Simple Case

8.21 The simple hedonic regression model defined by (8.2) will be estimated on this data set and is repeated here for convenience:

$$p_n^t = \beta^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t \qquad (8.9)$$
$$t = 1,...,14; \ n = 1,...,N(t)$$

The parameters to be estimated are β^t (i.e., the price of land in quarter t), γ^t (the price of constant quality structures in quarter t) and δ (the common depreciation rate for all quarters). Model (8.9) has 14 unknown β^t parameters, 14 unknown γ^t parameters and one unknown δ or 29 unknown parameters in all. [10]

8.22 The R^2 for this model was equal to .8847, which is the highest yet for regressions using the data set for the town of "A". The log likelihood was -10642.0, which is considerably higher than the log likelihoods for the two time dummy regressions that used prices as the dependent variable; recall the regression results associated with the construction of indices P_{H4} and P_{H5} defined in Chapter 5 where the log likelihoods were -10790.4 and -10697.8. The estimated decade straight line net depreciation rate was 0.1068 (0.00284).

8.23 The estimated land price series $\hat{\beta}^1,...,\hat{\beta}^{14}$ (rescaled to equal 1 in quarter 1), labeled P_{L1}, and quality adjusted price series for structures $\hat{\gamma}^1,...,\hat{\gamma}^{14}$ (rescaled also), labeled P_{S1}, are plotted in Figure 8.1 and listed in Table 8.1. Using these price series and the corresponding quantity data for each quarter t, i.e., the amount of land transacted, $L^t \equiv \sum_{n=1}^{N(t)} L_n^t$, and the quantity of constant quality structures, $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta} A_n^t) S_n^t$, an overall property price index has been constructed using the Fisher formula. This overall index, labeled P_1, is also plotted in Figure 8.1 and listed in Table 8.1. For comparison purposes, the Fisher hedonic imputation index from Chapter 5, P_{HIF}, is also presented.

---

[10] Model (8.9) is similar in structure to the hedonic imputation model described earlier except that the present model is more parsimonious; there is only one depreciation rate, as opposed to 14 depreciation rates in the imputation model defined by equations (5.25), and there is no constant term. The important factor in both models is that the prices of land and quality adjusted structures are allowed to vary independently across time periods.

Figure 8.1. The Price of Land (P$_{L1}$), the Price of Quality Adjusted Structures (P$_{S1}$), the Overall Cost of Production House Price Index (P$_1$) and the Fisher Hedonic Imputation House Price Index
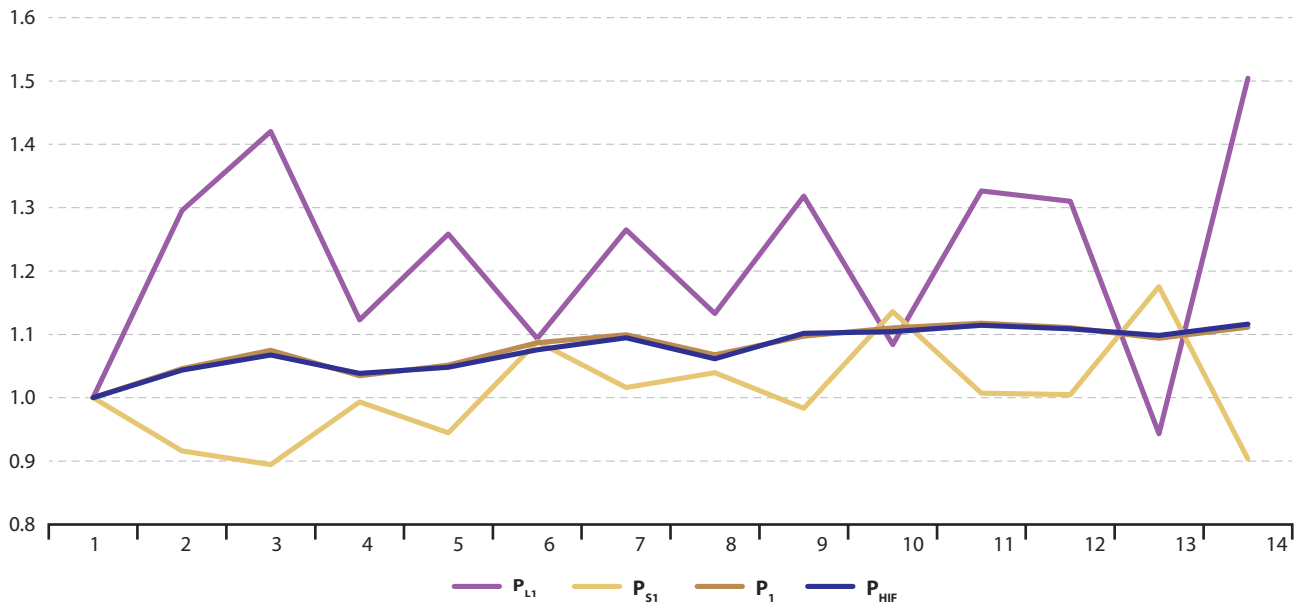
Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.1. The Price of Land (P$_{L1}$), the Price of Quality Adjusted Structures (P$_{S1}$), the Overall Cost of Production House Price Index (P$_1$) and the Fisher Hedonic Imputation House Price Index

Quarter	P$_{L1}$	P$_{S1}$	P$_1$	P$_{HIF}$
1	1.00000	1.00000	1.00000	1.00000
2	1.29547	0.91603	1.04571	1.04356
3	1.42030	0.89444	1.07482	1.06746
4	1.12290	0.99342	1.03483	1.03834
5	1.25820	0.94461	1.05147	1.04794
6	1.09346	1.08879	1.08670	1.07553
7	1.26514	1.01597	1.09941	1.09460
8	1.13276	1.03966	1.06787	1.06158
9	1.31816	0.98347	1.09713	1.10174
10	1.08366	1.13591	1.11006	1.10411
11	1.32624	1.00699	1.11782	1.11430
12	1.30994	1.00502	1.11077	1.10888
13	0.94311	1.17530	1.09373	1.09824
14	1.50445	0.9032	1.11147	1.11630

Source: Authors' calculations based on data from the Dutch Land Registry

8.24 It can be seen that the new overall hedonic price index based on a cost of production approach to the hedonic functional form, P_1, is very close to the Fisher hedonic imputation index P_{HIF}. However, the price series for land, P_{L1}, and the price series for quality adjusted structures, P_{S1}, are not credible at all: there are large random fluctuations in both series. Notice that when the price of land spikes upwards, there is a corresponding dip in the price of structures. This is a clear sign of multicollinearity between the land and quality adjusted structures variables, which leads to highly unstable estimates for the prices of land and structures.

The Use of Linear Splines

8.25 There is a tendency for the price of land per meter squared to decrease for large lots. In order to account for this, a *linear spline model for the price of land*

will be used.([11]) For lots that are less than 160 m², it is assumed that the cost of land per meter squared is β_S^t in quarter t. For properties that have lot sizes between 160 m² and 300 m², it is assumed that the cost of land changes to a price of β_M^t per additional square meter in quarter t. Finally, for plots above 300 m², the marginal price of an additional unit of land is set equal to β_L^t per square meter in quarter t. Let the sets of sales of small, medium and large plots be denoted by $S_S(t)$, $S_M(t)$ and $S_L(t)$, respectively, for $t = 1,...,14$. For sales n of properties that fall into the small land size group during quarter t, the hedonic regression model is given by (8.10); for the medium group by (8.11) and for the large land size group by (8.12):

$$p_n^t = \beta_S^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t \qquad (8.10)$$

$$t = 1,...,14; \; n \in S_S(t)$$

$$p_n^t = \beta_S^t [160] + \beta_M^t [L_n^t - 160] + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t \qquad (8.11)$$

$$t = 1,...,14; \; n \in S_S(t)$$

$$p_n^t = \beta_S^t [160] + \beta_M^t [140] + \beta_L^t [L_n^t - 300] + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t,$$

$$t = 1,...,14; \; n \in S_S(t) \qquad (8.12)$$

8.26 Estimating the model defined by (8.10)-(8.12) on the data for the town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1041$ (0.00419). The R² for this model was .8875, which is an increase over the previous no-splines model where the R² was .8847. The log likelihood was -10614.2 (an increase of 28 from the previous model's log likelihood.) The first period parameter values for the three marginal prices for land were $\hat{\beta}_S^1 = 281.4$ (55.9), $\hat{\beta}_M^1 = 380.4$ (48.5) and $\hat{\beta}_L^1 = 188.9$ (27.5). In other words, in quarter 1, the marginal cost per m² of small lots is estimated to be 281.4 Euros per m², for medium sized lots, the estimated marginal cost is 380.4 Euros/m², and for large lots, the estimated marginal cost is 188.9 Euros/m². The first period parameter value for quality adjusted structures is $\hat{\gamma}^1 = 978.1$ Euros/m² with a standard error of 82.3. The lowest t statistic for all of the 57 parameters was 3.3, so all of the estimated coefficients in this model are significantly different from zero.

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

([11]) This approach follows that of Diewert, de Haan and Hendriks (2010) (2011). The use of linear splines to model nonlinearities in the price of land as a function of lot size is due to Francke (2008).

$$V_{LS}^t \equiv \sum_{n \in S_S(t)} \hat{\beta}_S^t L_n^t \qquad (8.13)$$

$$t = 1,...,14$$

$$V_{LM}^t \equiv \sum_{n \in S_M(t)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [L_n^t - 160]\} \qquad (8.14)$$

$$t = 1,...,14$$

$$V_{LL}^t \equiv \sum_{n \in S_L(t)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [140] + \hat{\beta}_L^t [L_n^t - 300]\} \qquad (8.15)$$

$$t = 1,...,14$$

$$L_{LS}^t \equiv \sum_{n \in S_S(t)} L_n^t \qquad (8.16)$$

$$t = 1,...,14$$

$$L_{LM}^t \equiv \sum_{n \in S_M(t)} L_n^t \qquad (8.17)$$

$$t = 1,...,14$$

$$L_{LL}^t \equiv \sum_{n \in S_L(t)} L_n^t \qquad (8.18)$$

$$t = 1,...,14$$

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:

$$P_{LS}^t \equiv V_{LS}^t / L_{LS}^t; \; P_{LM}^t \equiv V_{LM}^t / L_{LM}^t; \; P_{LL}^t \equiv V_{LL}^t / L_{LL}^t \qquad (8.19)$$

$$t = 1,...,14$$

8.28 The average land prices for small, medium and large lots defined by equation (8.19) and the corresponding quantities of land defined by (8.16)-(8.18) can be used to construct a chained Fisher land price index, which is denoted by P_{L2}. This index is plotted in Figure 8.2 and listed in Table 8.2. As before, the estimated quarter t price per meter squared of quality adjusted structures is $\hat{\gamma}^t$ and the quantity of constant quality structures is given by $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta} A_n^t) S_n^t$. The structures price and quantity series $\hat{\gamma}^t$ 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 also graphed in Figure 8.2 and listed in Table 8.2. The constant quality structures price index P_{S2} (which is a normalization of the series $\hat{\gamma}^1,...,\hat{\gamma}^{14}$) is presented as well.

8.29 The overall house price index resulting from the spline model, P_2, is fairly close to the Fisher hedonic imputation index P_{HIF}. However, the spline model does not generate sensible series for the price of land, P_{L2}, and the price of structures, P_{S2}: both series are extremely volatile but in opposite directions. As was the case with the previous cost of production model, the present model suffers from a multicollinearity problem.

8.30 Comparing Figures 8.1 and 8.2, it can be seen that in Figure 8.1 the price index for land is above the overall price index for the most part and the price index for structures is below the overall index while in Figure 8.2, this pattern *reverses*. This instability is again an indication of multicollinearity. In the following section an attempt to cure this problem will be made by imposing monotonicity restrictions on the prices of the constant quality structures.

Figure 8.2. The Price of Land (P_{L2}), the Price of Structures (P_{S2}), the Overall Price Index Using Splines on Land (P_2) and the Fisher Hedonic Imputation Price Index
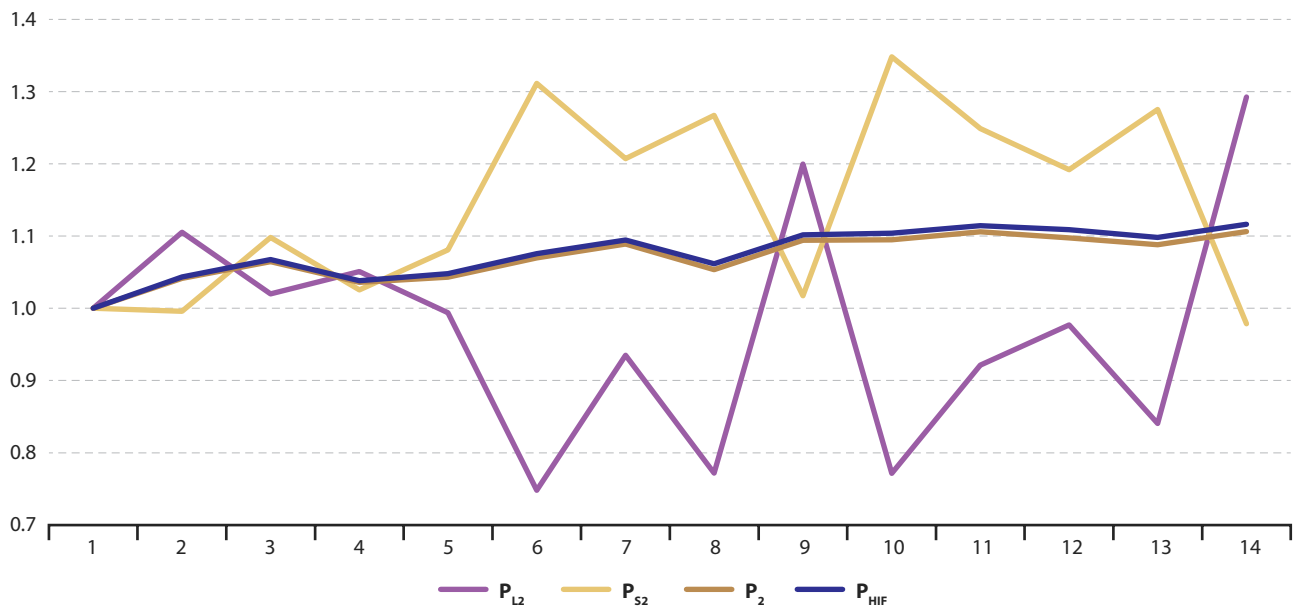
Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.2. The Price of Land (P_{L2}), the Price of Structures (P_{S2}), the Overall Price Index Using Splines on Land (P_2) and the Fisher Hedonic Imputation Price Index

Quarter	P_{L2}	P_{S2}	P_2	P_{HIF}
1	1.00000	1.00000	1.00000	1.00000
2	1.10534	0.99589	1.04137	1.04356
3	1.02008	1.09803	1.06465	1.06746
4	1.05082	1.02542	1.03608	1.03834
5	0.99379	1.08078	1.04294	1.04794
6	0.74826	1.31122	1.06982	1.07553
7	0.93484	1.20719	1.08912	1.09460
8	0.77202	1.26718	1.05345	1.06158
9	1.19966	1.01724	1.09425	1.10174
10	0.77139	1.34813	1.09472	1.10411
11	0.92119	1.24884	1.10596	1.11430
12	0.97695	1.19188	1.09731	1.10888
13	0.84055	1.27531	1.08811	1.09824
14	1.29261	0.97875	1.10613	1.11630

Source: Authors' calculations based on data from the Dutch Land Registry

An Approach Based on Monotonicity Restrictions

8.31 It is likely that Dutch construction costs did not fall significantly during the sample period.[12] If this is indeed the case, *monotonicity restrictions* on the quarterly prices of quality adjusted structures, $\gamma^1, \gamma^2, \gamma^3, ..., \gamma^{14}$, can be imposed on the hedonic regression model (8.10)-(8.12) by replacing the constant quality quarter t structures price parameters   by the following sequence of parameters for the 14 quarters: γ^1, $\gamma^1 + (\phi^2)^2$, $\gamma^1 + (\phi^2)^2 + (\phi^3)^2$,..., $\gamma^1 + (\phi^2)^2 + (\phi^3)^2 + ... + (\phi^{14})^2$, where $\phi^2, \phi^3, ..., \phi^{14}$ are scalar parameters.[13] For each quarter t starting at quarter 2, the price of a square meter of constant quality structures γ^t is thus equal to the previous period's price γ^{t-1} plus the square of a parameter ϕ^{t-1}, $(\phi^{t-1})^2$. Now replace this reparameterization of the structures price parameters γ^t in (8.10)-(8.12) in order to obtain a linear spline model for the price of land with monotonicity restrictions on the price of constant quality structures.

8.32 Implementing this new model using the data for the Dutch town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1031$ (0.00386). The R^2 for this model was .8859, a drop from the previous unrestricted spline model where the R^2 was .8875. The log likelihood was -10630.5,

a decrease of 16.3 over the previous unrestricted model. Eight of the 13 new parameters ϕ^t are zero in this monotonicity restricted hedonic regression. The first period parameter values for the three marginal land prices are $\hat{\beta}_S^1 = 278.6$ (37.2), $\hat{\beta}_M^1 = 380.3$ (41.0) and $\hat{\beta}_L^1 = 188.0$; these values are almost identical to the corresponding estimates in the previous unrestricted model. The first period parameter estimate for quality adjusted structures is $\hat{\gamma}^1 = 980.5$ (49.9) Euros/m2., which is little changed from the previous unrestricted estimate of 978.1 Euros/m2.

8.33 Once the parameters for the model have been estimated, convert the estimated ϕ^t parameters into estimated   parameters using the following recursive equations:

$$\hat{\gamma}^{t+1} \equiv \hat{\gamma}^t + (\hat{\phi}^t)^2 \qquad (8.19)$$

$$t = 2, ..., 14$$

Now use equations (8.13)-(8.19) in the previous section in order to construct a chained Fisher index of land prices, which is denoted by P_{L3}. This index is plotted in Figure 8.3 and listed in Table 8.3. As in the previous two models, the estimated period t price for a squared meter of quality adjusted structures is $\hat{\gamma}^t$ and the corresponding quantity of constant quality structures is $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta} A_n^t) S_n^t$. The price and quantity series $\hat{\gamma}^t$ and S^{t*} were combined with the three land price and quantity series to construct a chained overall Fisher house price index P_3 which is also graphed in Figure 8.3 and listed in Table 8.3. The constant quality structures price index P_{S3} (a normalization of the series $\hat{\gamma}^1, ..., \hat{\gamma}^{14}$) may be found in Figure 8.3 and Table 8.3 as well.

---

[12] Some direct evidence on this assertion will be presented in the following section.
[13] This method for imposing monotonicity restrictions was used by Diewert, de Haan and Hendriks (2010) with the difference that they imposed monotonicity on both structures and land prices, whereas here, monotonicity restrictions are imposed on structures prices only.

Figure 8.3. The Price of Land (P_{L3}), the Price of Quality Adjusted Structures (P_{S3}), the Overall House Price Index with Monotonicity Restrictions on Structures (P_3) and the Overall House Price Index Using Splines on Land (P_2)
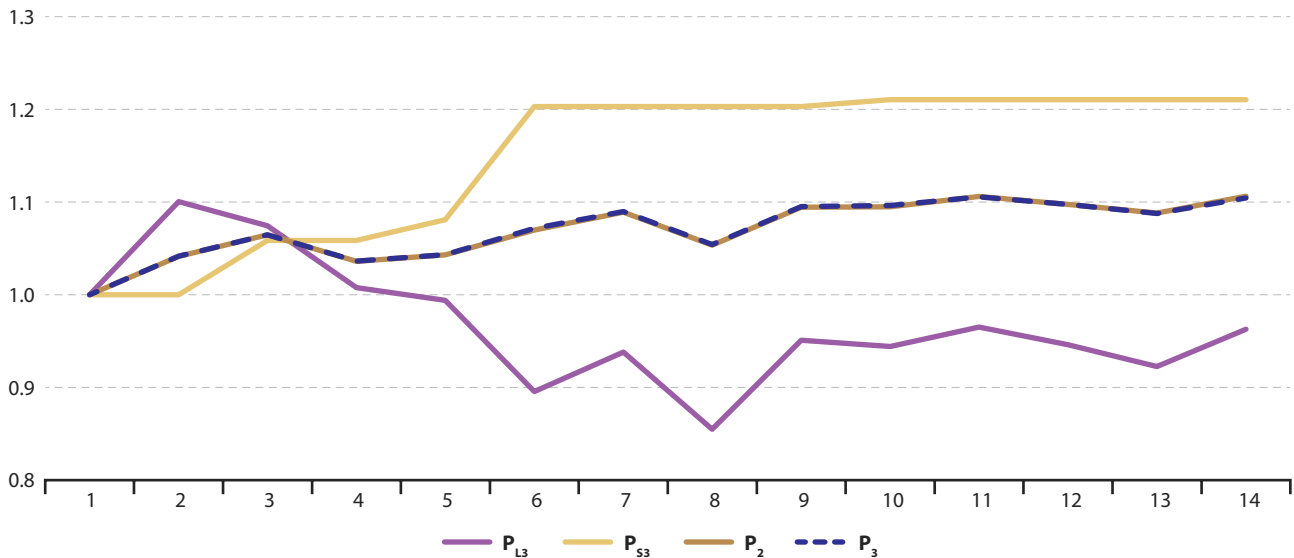
Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.3. The Price of Land (P_{L3}), the Price of Quality Adjusted Structures (P_{S3}), the Overall House Price Index with Monotonicity Restrictions on Structures (P_3) and the Overall House Price Index Using Splines on Land (P_2)

Quarter	P_{L3}	P_{S3}	P_3	P_2
1	1.00000	1.00000	1.00000	1.00000
2	1.10047	1.00000	1.04148	1.04137
3	1.07431	1.05849	1.06457	1.06465
4	1.00752	1.05849	1.03627	1.03608
5	0.99388	1.08078	1.04316	1.04294
6	0.89560	1.20300	1.07168	1.06982
7	0.93814	1.20300	1.08961	1.08912
8	0.85490	1.20300	1.05408	1.05345
9	0.95097	1.20300	1.09503	1.09425
10	0.94424	1.21031	1.09625	1.09472
11	0.96514	1.21031	1.10552	1.10596
12	0.94596	1.21031	1.09734	1.09731
13	0.92252	1.21031	1.08752	1.08811
14	0.96262	1.21031	1.10427	1.10613

Source: Authors' calculations based on data from the Dutch Land Registry

8.34 The new overall house price index P_3 that imposed monotonicity on the quality adjusted price of structures in Figure 8.3 can hardly be distinguished from the previous overall house price index P_2, which was based on a similar hedonic regression model except that the movements in the price of structures were not restricted. The fluctuations in the price of land and quality adjusted structures are no longer violent.

8.35 While the above results seem "reasonable", the early rapid rise in the price of structures and the slow growth in structures prices from quarter 6 to 14 are not very likely. In the following section, one more method for extracting separate structures and land components out of real estate sales data will therefore be tried.

An Approach Based on Exogenous Information on the Price of Structures

8.36 Many countries have new construction price indices available on a quarterly basis. This is the case for the Netherlands.[14] If one is willing to make the assumption that construction costs for houses have the same rate of growth over the study period across all cities in the Netherlands, the information on construction costs can be used to eliminate the multicollinearity problem encountered in the previous sections.

8.37 Recall equations (8.10)-(8.12) above. These are the estimating equations for the unrestricted hedonic regression model based on costs of production. In the present section, the constant quality price parameters for the structures, the γ^t for $t = 2,...,14$ in (8.10)-(8.12), are replaced by the following numbers, which involve only the single unknown parameter γ^1:[15]

$$\gamma^t = \gamma^1 \mu^t \qquad (8.20)$$

$$t = 2,...,14$$

where μ^t is the statistical agency's *construction cost price index* for the location and the type of house under consideration, normalized to equal 1 in quarter 1. The new hedonic regression model is again defined by equations (8.10)-(8.12) except that the 14 unknown γ^t parameters are now defined by (8.20), so that only γ^1 needs to be estimated. The number of parameters to be estimated in this new restricted model is 44 whereas the old number was 57.

8.38 Using the data for the town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1028$ (0.00433). The R^2 for this model was .8849, a small drop from the previous restricted spline model, where the R^2 was .8859, and a larger drop from the unrestricted spline model R^2 in section 8.5, which was .8875. The log likelihood was -10640.1,

[14] From the Statistics Netherlands (2010) online source, Statline, the following series was downloaded for the New Dwellings Output Price Index for the 14 quarters in our sample of house sales: 98.8, 98.1, 100.3, 102.7, 99.5, 100.5, 100.0, 100.3, 102.2, 103.2, 105.6, 107.9, 110.0, 110.0. This series was normalized to 1 in the first quarter by dividing each entry by 98.8. The resulting series is denoted by μ^1 (=1), μ^2,...,μ^{14}.

[15] The technique suggested here for decomposing property prices into land and structures components can be viewed as a variant of a technique used by Davis and Heathcote (2007) and Davis and Palumbo (2008).

a decrease of 10 over the monotonicity restricted model. The first period parameter estimates for the 3 marginal prices for land are now $\hat{\beta}_S^1 = 215.4$ (30.0), $\hat{\beta}_M^1 = 362.6$ (46.7) and $\hat{\beta}_L^1 = 176.4$ (28.4). They differ slightly from the previous figures. The first period parameter estimate for the quality adjusted structures is $\hat{\gamma}^1 = 1085.9$ (22.9) Euros/m², which is significantly higher than the unrestricted estimate of 980.5 Euros/m². So the imposition of a (nationwide) growth rate on the change in the price of quality adjusted structures has had some effect on the estimates for the levels of land and structures prices.

8.39 As usual, equations (8.13)-(8.19) were used in order to construct a chained Fisher index of land prices, which is denoted by P_{L4}. This index is plotted in Figure 8.4 and listed in Table 8.4. As for the previous three models, the estimated price in quarter t for a square meter of quality adjusted structures is $\hat{\gamma}^t$ (which now equals $\hat{\gamma}^1 \hat{\mu}^t$) and the corresponding quantity is $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta}A_n^t)S_n^t$. These structures price and quantity series were again

combined with the three land price and quantity series to form a chained overall Fisher house price index P_4, which is graphed in Figure 8.4 and listed in Table 8.4. The constant quality structures price index P_{S4} (a normalization of the series $\hat{\gamma}^1, ..., \hat{\gamma}^{14}$) is also presented.

8.40 A comparison of Figures 8.3 and 8.4 shows that the imposition of the national growth rates for new dwelling construction costs has changed the nature of the land and structures price indices: in Figure 8.3, the price series for land lies below the overall house price series for most of the sample period while in Figure 8.4, the pattern is reversed: the price series for land lies above the overall house price series for most of the sample period (and vice versa for the price of structures). But which model is best? Although the previous model can be preferred on statistical grounds because the log likelihood is somewhat higher, we would nevertheless prefer the present model that uses of exogenous information on structures prices because it yields a more plausible pattern of price changes for land and structures.

Figure 8.4. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}) and the Overall House Price Index using Exogenous Information on the Price of Structures (P_4)
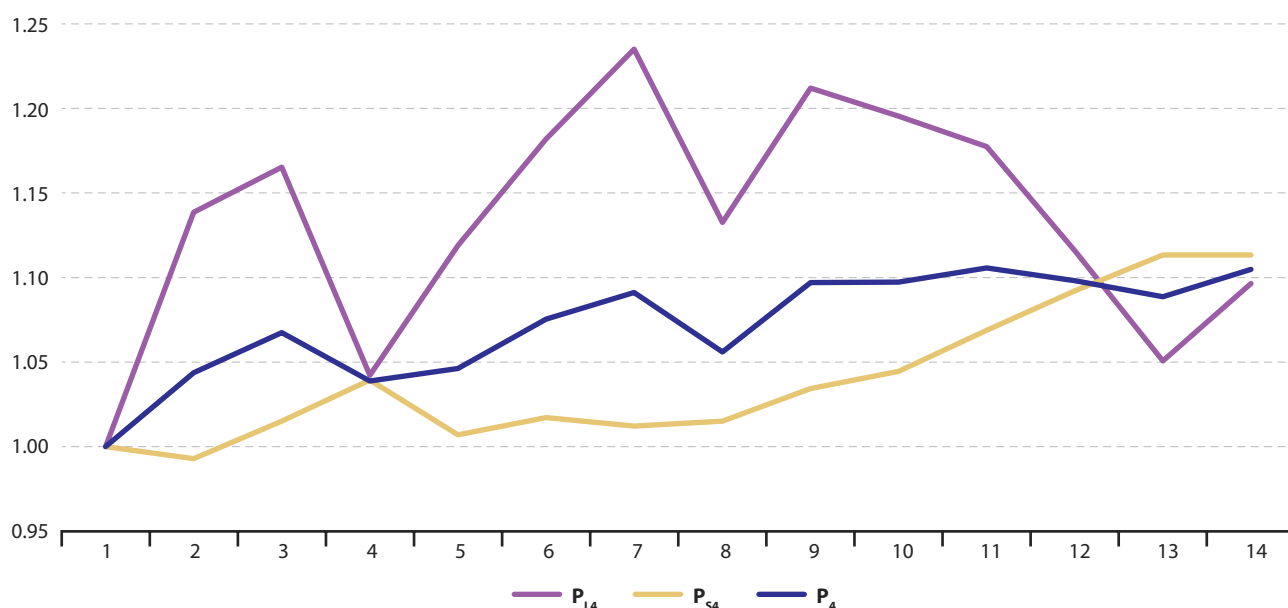
Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.4. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}) and the Overall House Price Index using Exogenous Information on the Price of Structures (P_4)

Quarter	P_{L4}	P_{S4}	P_4
1	1.00000	1.00000	1.00000
2	1.13864	0.99291	1.04373
3	1.16526	1.01518	1.06752
4	1.04214	1.03947	1.03889
5	1.11893	1.00709	1.04628
6	1.18183	1.01721	1.07541
7	1.23501	1.01215	1.09121
8	1.13257	1.01518	1.05601
9	1.21204	1.03441	1.09701
10	1.19545	1.04453	1.09727
11	1.17747	1.06883	1.10564
12	1.11588	1.09211	1.09815
13	1.05070	1.11336	1.08863
14	1.09648	1.11336	1.10486

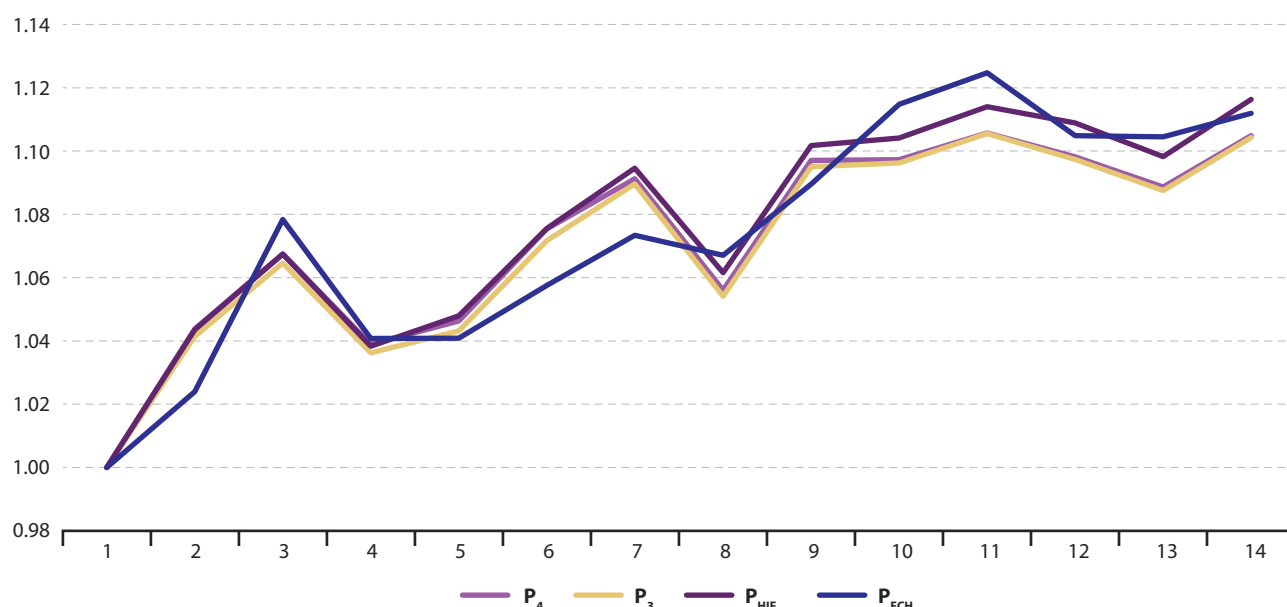*Source:* Authors' calculations based on data from the Dutch Land Registry

Choosing the "Best" Overall Index

8.41 This section is concluded by listing and charting our four "best" *overall* indices: the chained stratified sample Fisher index P_{FCH} constructed in Chapter 4, the chained hedonic imputation Fisher index P_{HIF} studied in Chapter 5, the index P_3 that resulted from the cost based hedonic regression model with monotonicity restrictions constructed earlier, and the index P_4 that resulted from the cost based hedonic regression model using exogenous information on the price of structures studied in the present section. As can be seen from Figure 8.5, all four indices paint much the same picture. Note that P_3 and P_4 are virtually identical.

8.42 All things considered, the hedonic imputation index P_{HIF} is our preferred index since it has fewer restrictions than the other indices and seems closest to a matched model index in spirit, followed by the two cost of production hedonic indices P_4 and P_3, followed by the stratified sample index P_{FCH}. The latter likely suffers from some unit value bias. Hedonic indices can be biased too (if important explanatory variables are omitted or if an "incorrect" functional form is chosen), but in general we would prefer hedonic regression methods over stratification methods. If separate land and structures indices are required, we are in favour of the cost based hedonic regression model that uses exogenous information on the price of structures.

Figure 8.5. House Price Indices Using Exogenous Information (P$_4$) and Using Monotonicity Restrictions (P$_3$), the Chained Fisher Hedonic Imputation Index and the Chained Fisher Stratified Sample Index

Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.5. House Price Indices Using Exogenous Information (P$_4$) and Using Monotonicity Restrictions (P$_3$), the Chained Fisher Hedonic Imputation Index and the Chained Fisher Stratified Sample Index

Quarter	P$_4$	P$_3$	P$_{HIF}$	P$_{FCH}$
1	1.00000	1.00000	1.00000	1.00000
2	1.04373	1.04148	1.04356	1.02396
3	1.06752	1.06457	1.06746	1.07840
4	1.03889	1.03627	1.03834	1.04081
5	1.04628	1.04316	1.04794	1.04083
6	1.07541	1.07168	1.07553	1.05754
7	1.09121	1.08961	1.09460	1.07340
8	1.05601	1.05408	1.06158	1.06706
9	1.09701	1.09503	1.10174	1.08950
10	1.09727	1.09625	1.10411	1.11476
11	1.10564	1.10552	1.11400	1.12471
12	1.09815	1.09734	1.10888	1.10483
13	1.08863	1.08752	1.09824	1.10450
14	1.10486	1.10427	1.11630	1.11189

Source: Authors' calculations based on data from the Dutch Land Registry

Rolling Window Hedonic Regressions

8.43 A problem with the hedonic regression model discussed in the previous section (and all other hedonic models discussed in this Handbook except hedonic imputation models) was mentioned in Chapter 5: when more data are added, the indices generated by the model change. This feature of these regression based methods makes these models unsatisfactory for statistical agency use, where users expect the official numbers to remain unchanged as time passes. Users may tolerate a few revisions to recent data but typically, they would not like all the numbers to be revised back into the indefinite past as new data become available. A simple solution to this problem is available, however. the so-called *rolling window approach*. This approach will be outlined in more detail and applied to the cost based hedonic regression model that uses exogenous information on the price of structures.

8.44 First, one chooses a "suitable" number of time periods (equal to or greater than two) where it is thought that the hedonic model yields "reasonable" results; this will be the *window length* (say M periods) for the sequence of regression models which will be estimated. Secondly, an initial regression model is estimated and the appropriate indices 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 indices 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*; for a general window length, it is called a *rolling window hedonic regression model*.[16]

8.45 Using the data for the town of "A", the rolling window procedure was applied with a window length of 9 quarters. The hedonic regression model defined by equations (8.10)-(8.12) and (8.20) was initially estimated for the first 9 quarters. The resulting price indices for land and for constant quality structures and the overall index are denoted by P_{RWL4}, P_{RWS4} and P_{RW4} and are listed in the first 9 rows of Table 8.6.[17] Next, a regression covering quarters 2-10 was run and the resulting land, structures and overall price indices were used to update the initial indices; i.e., the price of land in quarter 10 of Table 8.6 is equal to the price of land in quarter 9 times the price relative for land (quarter 10 land index divided by the quarter 9 land index) obtained from the regression covering quarters 2-10, etc. Similar updating was done for the next 4 quarters using regressions covering quarters 3-11, 4-12, 5-13 and 6-14.

8.46 The rolling window indices can be compared to the corresponding indices based on the data pertaining to all 14 quarters constructed in the previous section by looking at Table 8.6. Recall that the estimated depreciation rate and the estimated quarter 1 price of quality adjusted structures for the last model were $\hat{\delta} = 0.1028$ and $\hat{\gamma}^1 = 1085.9$, respectively. If by chance the 6 rolling window hedonic regressions generated the exact same estimates for δ and γ, then the indices resulting from the rolling window regressions would coincide with the indices P_{L4}, P_{S4} and P_4. The estimates for δ generated by the 6 rolling window regressions are 0.10124, 0.10805, 0.11601, 0.11103, 0.10857 and 0.10592. The estimates for γ^1 generated by the 6 rolling window regressions are 1089.6, 1103.9, 1088.1, 1101.0, 1123.5 and 1100.9. While these estimates are not identical to the corresponding estimates of 0.1028 and 1085.9 for P_4, they are fairly close. So we can expect the rolling window indices to be close to their counterparts for the last model in the previous section. The R^2 values for the 6 rolling window regressions were .8803, .8813, .8825, .8852, .8811 and .8892.

8.47 The rolling window series for the price of quality adjusted structures, P_{RWS}, is not listed in Table 8.6 since it is identical to the series P_{S4}.[18] The rolling window price series for land, P_{RWL}, is extremely close to its counterpart P_{L4}, and the overall rolling window price series for detached dwellings in the town of "A", P_{RW}, is also close to its counterpart P_4. The corresponding series in Table 8.6 are so close to each other that we decided not to provide a chart.

---

[16] This procedure was recently used by Shimizu, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010) in their hedonic regression models for Tokyo house prices. An analogous procedure has also been recently applied by Ivancic, Diewert and Fox (2011) and de Haan and van der Grient (2011) in their adaptation of the GEKS method for making international comparisons to the scanner data context.

[17] We imposed the restrictions (33) on the rolling window regressions and so the rolling window constant quality price index for structures, P_{RWS}, is equal to the constant quality price index for structures listed in Table 8.4, P_{S4}.

[18] By construction, P_{S4} and P_{RWS} are both equal to the official Statistics Netherlands construction price index for new dwellings, μ^t/μ^1 for t = 1,...,14.

Table 8.6. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}), the Overall House Price Index using Exogenous Information on the Price of Structures (P_4) and their Rolling Window Counterparts (P_{RWL}) and (P_{RW})

Quarter	P_{RWL}	P_{L4}	P_{RW}	P_4	P_{S4}
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.14073	1.13864	1.04381	1.04373	0.99291
3	1.16756	1.16526	1.06766	1.06752	1.01518
4	1.04280	1.04214	1.03909	1.03889	1.03947
5	1.12055	1.11893	1.04635	1.04628	1.00709
6	1.18392	1.18183	1.07542	1.07541	1.01721
7	1.23783	1.23501	1.09123	1.09121	1.01215
8	1.13408	1.13257	1.05602	1.05601	1.01518
9	1.21417	1.21204	1.09698	1.09701	1.03441
10	1.19772	1.19545	1.09738	1.09727	1.04453
11	1.18523	1.17747	1.10718	1.10564	1.06882
12	1.11889	1.11588	1.09779	1.09815	1.09201
13	1.05191	1.05070	1.08893	1.08863	1.11335
14	1.09605	1.09648	1.10436	1.10486	1.11335

Source: Authors' calculations based on data from the Dutch Land Registry

8.48 Using the data for the town of "A", rolling window hedonic regressions gave much the same results as a hedonic regression that covers the whole sample period. This supports our view that the rolling window approach can be used by statistical agencies to compile an RPPI based on hedonic regressions, including a decomposition into land and structures components.

The Construction of Price Indices for the Stock of Dwelling Units

8.49 This section shows how hedonic regression models can be used to form an approximate RPPI for the *stock* of dwelling units. We will first look at the hedonic imputation model discussed in Chapter 5 and compare the resulting index with an approximate stock based index using the stratification approach.

The Hedonic Imputation Model

8.50 Recall that the hedonic imputation model was defined by equations (5.25), where L_n^t, S_n^t and A_n^t denoted, respectively, the land area, structure area, and age (in decades) of property n sold in period t. To form a price index for the stock of detached houses in the town of "A", it would in principle be necessary to know L, S and A for all detached houses in "A" during some base period. This

information is not available to us, but we can treat the total number of detached houses sold over the sample period as an *approximation* to the stock of this type.[19] In our data set there were $N(1) + N(2) + ... + N(14) = 2289$ of such transactions.[20]

8.51 The estimated parameters for land size, structure size and depreciation in quarter t are denoted by $\hat{\beta}^t$, $\hat{\gamma}^t$ and $\hat{\delta}^t$; $\hat{\alpha}^t$ denotes the constant term. Our approximation to *the total value of the housing* stock for quarter t, V^t, is defined as

$$V^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} [\hat{\alpha}^t + \hat{\beta}^t L_n^s + \hat{\gamma}^t (1 - \hat{\delta}^t A_n^s) S_n^s] \qquad (8.21)$$

$$t = 1,...,14$$

That is, V^t is (approximated by) the imputed value of all houses traded during the 14 quarters in our sample, where the regression coefficients from the quarter t hedonic imputation model given by (5.25) serve as weights for the characteristics of each house. Dividing the V^t series by the value for quarter 1, V^1, is our first estimated stock price index, P_{Stock1}, for the town of "A".[21] This is a form of a Lowe index; see the CPI Manual (2004) for the properties

[19] This approximation would probably be an adequate one if the sample period were a decade or so. Obviously, our sample period of 14 quarters is too short to be accurate and there are also sample selectivity problems, i.e., newer houses will be over represented. However, the method we are suggesting here can be illustrated using this rough approximation.

[20] We did not delete the observations for houses that were transacted multiple times over the 14 quarters since a particular house transacted during two or more of the quarters is not actually the same house due to depreciation and renovations.

[21] Since V^t is a value, it does not appear to be a price series at first glance. But in each quarter, the quantity vector which underlies this value is a vector of ones of dimension 2289, which is constant over the 14 quarters. Hence V^t can also be interpreted as a price series, which is normalized to equal one in quarter 1.

of Lowe indices. In Table 8.7 and Figure 8.6 this price index for the stock of houses is compared with the corresponding sales based Fisher hedonic imputation price index, P_{HIF}.

8.52 An additional approximate stock price index based on *stratification*, P_{Stock2} is also graphed in Figure 8.6 and listed in Table 8.7. This index uses the unit value prices for the nonempty cells in the stratification scheme in each quarter, as explained in Chapter 4, and uses the imputed prices based on the hedonic imputation regressions from Chapter 5 for the empty cells in each quarter. The quantity vector used for P_{Stock2} is the (sample) total quantity vector by cell, which makes P_{Stock2} an alternative Lowe price index. It can be seen that while P_{Stock2} has the same general trend as P_{Stock1} and P_{HIF}, it differs substantially from these

hedonic imputation indices during several quarters. These differences are due to the existence of some unit value bias in the stratification indices. Thus, although stratification indices can be constructed for the stock of dwelling units of a certain type and location (with the help of hedonic imputation for empty cells), it appears that the resulting stock indices will not be as accurate as indices that are entirely based on the use of hedonic regressions. [22]

---

[22] If the imputed prices are used for every one of the 45 cell prices for each period (instead of just for the zero transaction cells as was the case for the construction of $P_{Sctock2}$) and the same total sample quantity vector is used as the approximate stock quantity vector, then the resulting Lowe index turns out to be exactly equal to P_{Stock1}. Thus these two different ways for constructing a stock index turn out to be equivalent. The fact that P_{Stock1} is not equal to P_{Stock2} is clear evidence that there is unit value bias in the cells of the stratification scheme: the cells are simply not defined narrowly enough.

Figure 8.6. Approximate Stock Price Indices and Based on Hedonic Imputation (P$_{Stock1}$) and Stratification (P$_{Stock2}$) and the Fisher Hedonic Imputation Sales Price Index
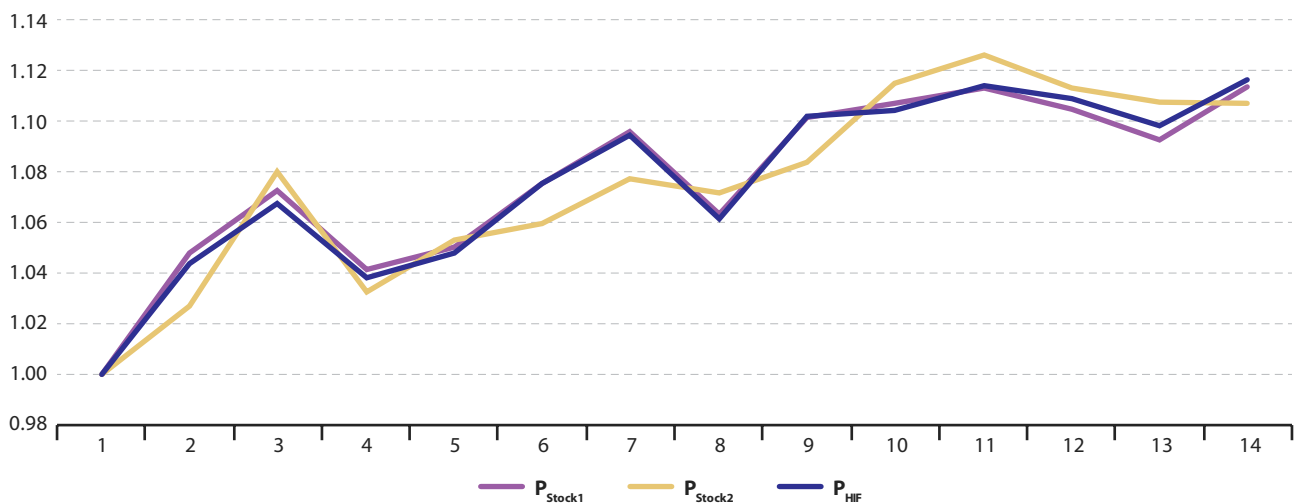
Source: Authors' calculations based on data from the Dutch Land Registry

Table 8.7. Approximate Stock Price Indices and Based on Hedonic Imputation (P$_{Stock1}$) and Stratification (P$_{Stock2}$) and the Fisher Hedonic Imputation Sales Price Index

Quarter	P$_{Stock1}$	P$_{Stock2}$	P$_{HIF}$
1	1.00000	1.00000	1.00000
2	1.04791	1.02712	1.04356
3	1.07255	1.07986	1.06746
4	1.04131	1.03257	1.03834
5	1.05040	1.05290	1.04794
6	1.07549	1.05934	1.07553
7	1.09594	1.07712	1.09460
8	1.06316	1.07172	1.06158
9	1.10137	1.08359	1.10174
10	1.10708	1.11482	1.10411
11	1.11289	1.12616	1.11430
12	1.10462	1.11291	1.10888
13	1.09278	1.10764	1.09824
14	1.11370	1.10686	1.11630

Source: Authors' calculations based on data from the Dutch Land Registry

The Use of Exogenous Information on the Price of Structures

8.53 The same kind of construction of an approximate stock price index can be applied to the other hedonic regression models discussed in this chapter. Here we will show how this works for the cost based model that used exogenous information on the price of structures. This model was defined by equations (8.10)-(8.12) and (8.20). Recall that the sets of period t sales of small, medium and large lot houses were denoted by $S_S(t)$, $S_M(t)$ and $S_L(t)$, respectively; the total number of sales in period t was denoted by $N(t)$ for $t = 1,...,14$. The estimated model parameters are $\hat{\delta}^t$, $\hat{\gamma}^t$ and $\hat{\beta}_S^1$, $\hat{\beta}_M^1$ and $\hat{\beta}_L^1$ for $t = 1,...,14$. The estimated *period t values of all small, medium and large lot houses* traded over the 14 quarters, V_{LS}^t, V_{LM}^t and V_{LL}^t, respectively, are defined by (8.22)-(8.24):

$$V_{LS}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_S(s)} \hat{\beta}_S^t L_n^s \qquad (8.22)$$

$$t = 1,...,14$$

$$V_{LM}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_M(s)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [L_n^s - 160]\} \qquad (8.23)$$

$$t = 1,...,14$$

$$V_{LL}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_L(s)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [140] + \hat{\beta}_L^t [L_n^s - 300]\} \qquad (8.24)$$

$$t = 1,...,14$$

The estimated period t value of quality adjusted structures, V_S^t, is defined by

$$V_S^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} \hat{\gamma}^1 \mu^t (1 - \hat{\delta} A_n^s) S_n^s \qquad (8.25)$$

$$t = 1,...,14$$

where all structures traded during the 14 quarters are included.

8.54 The *quantities* that correspond to the above period t valuations of the three land stocks and the stock of structures are defined as follows: [23]

$$Q_{LS}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_S(s)} L_n^s \qquad (8.26)$$

$$t = 1,...,14$$

$$Q_{LM}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_M(s)} L_n^s \qquad (8.27)$$

$$t = 1,...,14$$

$$Q_{LL}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_L(s)} L_n^s \qquad (8.28)$$

$$t = 1,...,14$$

$$Q_S^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} (1 - \hat{\delta} A_n^s) S_n^s \qquad (8.29)$$

$$t = 1,...,14$$

[23] The quantities defined by (8.26)-(8.29), which are constant over the 14 quarters, are equal to 77455, 258550, 253590 and 238476 for small lots, medium size lots, large lots and structures, respectively.

8.55 *Approximate stock prices*, P_{LS}^t, P_{LM}^t, P_{LL}^t and P_S^t, that correspond to the values and quantities defined by (8.22)-(8.29), can be computed in the usual way:

$$P_{LS}^t \equiv V_{LS}^t / Q_{LS}^t \qquad (8.30)$$

$$P_{LM}^t \equiv V_{LM}^t / Q_{LM}^t$$

$$P_{LL}^t \equiv V_{LL}^t / Q_{LL}^t$$

$$P_S^t \equiv V_S^t / Q_S^t$$

$$t = 1,...,14$$

Using the above prices and quantities, an approximate stock index of land prices, P_{LStock}, is formed by aggregating the three types of land and an approximate constant quality stock price index for structures, P_{SStock}, is simply formed by normalizing the series P_S^t. The approximate

overall stock index, P_{Stock}, is obtained by aggregating the three types of land with the constant quality structures (or, equivalently, by aggregating P_{LStock} and P_{SStock}). Since the quantities are constant over all 14 quarters, the Laspeyres, Paasche and Fisher price indices are all equal.[24] The stock price indices P_{LStock}, P_{SStock} and P_{Stock} are charted in Figure 8.7 and listed in Table 8.8. For comparison purposes, the corresponding price indices based on *sales* of properties for the model presented previously, P_{L4}, P_{S4} and P_4, are also listed in Table 8.8. As can be seen from Table 8.8, the approximate stock price index for structures P_{SStock} coincides with the sales based price index for constant quality structures P_{S4}, so P_{S4} is not charted in Figure 8.7.

[24] Fixed base and chained Laspeyres, Paasche and Fisher indices are also equal under these circumstances.

Figure 8.7. Approximate Price Indices for the Stock of Houses (P_Stock), the Stock of Land (P_LStock), the Stock of Structures (P_SStock) and the Corresponding Sales Indices (P_L4 and P_4)
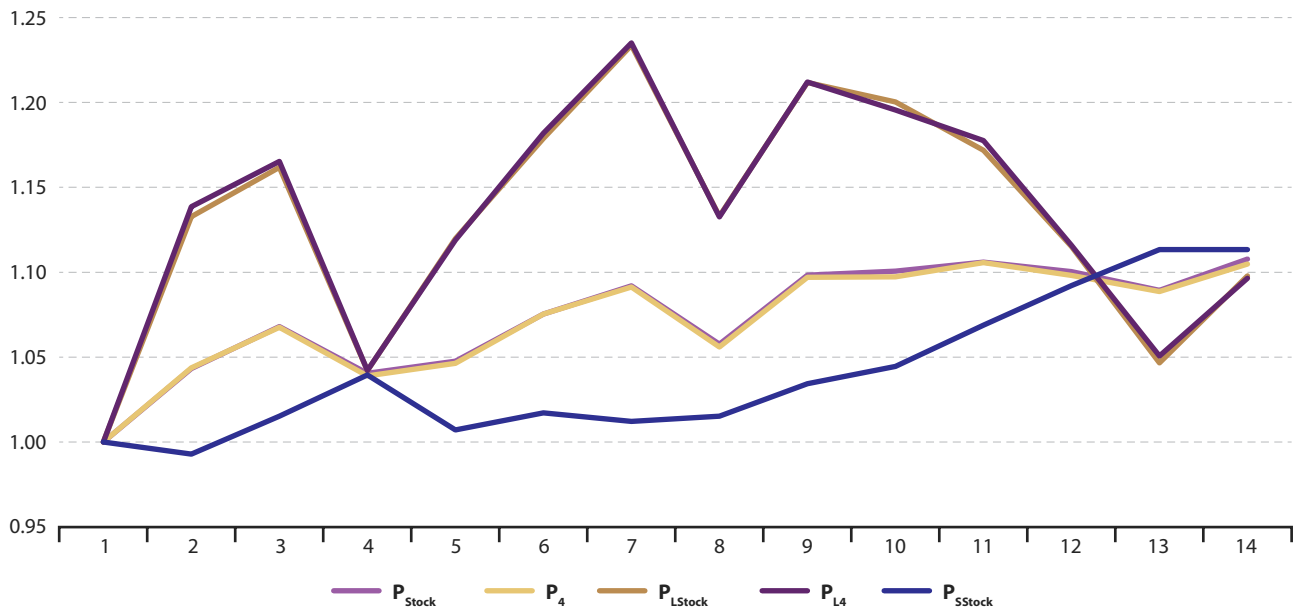
*Source:* Authors' calculations based on data from the Dutch Land Registry

Table 8.8. Approximate Price Indices for the Stock of Houses (P$_{Stock}$), the Stock of Land (P$_{LStock}$), the Stock of Structures (P$_{SStock}$) and the Corresponding Sales Indices (P$_{L4}$ and P$_4$)

Quarter	P$_{Stock}$	P$_4$	P$_{LStock}$	P$_{L4}$	P$_{SStock}$	P$_{S4}$
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.04331	1.04373	1.13279	1.13864	0.99291	0.99291
3	1.06798	1.06752	1.16171	1.16526	1.01518	1.01518
4	1.04042	1.03889	1.04209	1.04214	1.03947	1.03947
5	1.04767	1.04628	1.11973	1.11893	1.00709	1.00709
6	1.07540	1.07541	1.17873	1.18183	1.01721	1.01721
7	1.09192	1.09121	1.23357	1.23501	1.01215	1.01215
8	1.05763	1.05601	1.13299	1.13257	1.01518	1.01518
9	1.09829	1.09701	1.21171	1.21204	1.03441	1.03441
10	1.10065	1.09727	1.20029	1.19545	1.04453	1.04453
11	1.10592	1.10564	1.17178	1.17747	1.06883	1.06883
12	1.10038	1.09815	1.11507	1.11588	1.09211	1.09211
13	1.08934	1.08863	1.04668	1.05070	1.11336	1.11336
14	1.10777	1.10486	1.09784	1.09648	1.11336	1.11336

Source: Authors' calculations based on data from the Dutch Land Registry

8.56 The overall approximate price index for the total stock of detached houses in the town of "A" (P_{Stock}) can hardly be distinguished from the corresponding overall sales price index (P_4) in Figure 8.7. Similarly, the approximate price index for the stock of land in "A" (P_{LStock}) can barely be distinguished in Figure 8.7 from the corresponding sales price index for land (P_{L4}). Nevertheless, there are small differences between the stock and sales indices, as Table 8.8 shows.

8.57 Our conclusion is that the hedonic regression models for the sales of houses can readily be adapted to compute Lowe type price indices for the stock of houses. There do not appear to be major differences between the two index types when using our data set, but this result may not hold for other data sets.
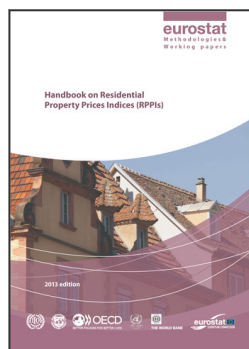
From:

Handbook on Residential Property Price Indices

OECD