Alternative Approaches to Measuring House Price Inflation

W. Erwin Diewert,[1]
Discussion Paper 10-10,
Department of Economics,
The University of British Columbia,
Vancouver, Canada, V6T 1Z1.
email: diewert@econ.ubc.ca

December 27, 2010
Revised: April 10, 2011

Abstract

The paper uses data on sales of detached houses in a small Dutch town over 14 quarters starting at the first quarter of 2005 in order to compare various methods for constructing a house price index over this period. Four classes of methods are considered: (i) stratification techniques plus normal index number theory; (ii) time dummy hedonic regression models; (iii) hedonic imputation techniques and (iv) additive in land and structures hedonic regression models. The last approach is used in order to decompose the price of a house into land and structure components and it relies on the imposition of some monotonicity constraints or exogenous information on price movements for structures. The problems associated with constructing an index for the stock of houses using information on the sales of houses are also considered.

Key Words

Property price indexes, hedonic regressions, stratification techniques, rolling year indexes, 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, 2010, December 3, Crowne Plaza Hotel, Coogee Beach, Sydney, Australia. Revised February 1, 2011. W. Erwin Diewert: Department of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and School of Economics, University of New South Wales. E-mail: diewert@econ.ubc.ca. The author thanks Ken Clements, Jan de Haan, Ronald Johnson, Ulrich Kohli, Alice Nakamura and James Morley for helpful comments. The author gratefully acknowledges 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 content of this paper.

Alternative Approaches to Measuring House Price Inflation

1. Introduction

This paper has two main purposes:

- Some real estate data for sales of detached houses in the Dutch town of "A" is used in order to construct house price indexes using a variety of methods. A main purpose of the paper is to determine whether the different methods generate different empirical results. The data cover 14 quarters of sales, beginning in 2005 and ending in the middle of 2008.
- The second main purpose is to determine whether it is possible to decompose an overall house price index into reliable Land and Structures components. This decomposition is required for some national income accounting purposes, as well as being of general interest.

With respect to the second main purpose, the present paper is a follow up on Diewert, Haan and Hendriks (2010). Those authors used a hedonic regression approach to decompose an overall house price index into land and structures components. Their decomposition method relied on the imposition of monotonicity restrictions on the prices of the two components and their approach worked satisfactorily because during the time period they studied, house prices in the Dutch town of "A" only rise. However, during the time period used in the present paper, house prices in the town of "A" both rise and fall and thus the methodology used by Diewert, Haan and Hendriks needs to be modified in order to deal with this problem.

With respect to the comparison of methods purpose, *four main classes of methods* for constructing house price indexes for *sales* of properties will be considered:

- *Stratification methods*; i.e., sales of houses during a period are segmented into relatively homogeneous classes and normal index number theory is applied to the cell data;
- *Time dummy hedonic regression methods*;
- *Hedonic regression imputation methods*;
- *Additive hedonic regression methods* with the imposition of period to period monotonicity restrictions to smooth the estimates for the land and structure components of the overall index.

The last three classes of methods are all variants of hedonic regressions.[2] The additive method is a variant of the method that was used by Diewert, Haan and Hendriks (2010).

[2] The difference between time dummy and imputation hedonic regressions has been theoretically analysed by Diewert, Heravi and Silver (2009) and Haan (2009) (2010).

All four classes of methods can be given theoretical justifications so it is of some interest to see how different or similar they are when implemented on the same data set.

A brief outline of the contents of each section follows.

In section 2, *stratification methods* are explained along with our data on real estate transactions for the small Dutch town of "A" over a 14 quarter period. This same data set will be used to illustrate how all of the various methods for constructing house price indexes work in practice.

The results from section 2 indicate that prices may follow a *seasonal* pattern of decline in the fourth quarter of each year. Solutions to this seasonality problem are explained in section 3.

In sections 4 and 5, standard hedonic regressions are implemented on the data set. There are three main characteristics of a detached house that sold in a quarter that are used in the hedonic regression: the age A of the house, its structure floor space area S and the land area of the plot L. The use of just these three characteristics leads to a hedonic regression that explain 84 to 89% of the variation in selling prices. In section 4, the dependent variable is the logarithm of the selling price while in section 5, we study hedonic regressions that use just the selling price as the dependent variable. The regressions in these two sections use the *time dummy methodology*.

In section 6, the time dummy methodology is not used. Instead, a separate hedonic regression for the data of each quarter is estimated and then these regressions were used to create imputed prices for the various "models" of houses that transacted so that a matched model methodology can be applied. This class of methods for constructing a house price index is based on what is called the *hedonic imputation methodology*. This method turns out to be our preferred method for constructing an overall house price index.

In section 7, we turn our attention to the problem of constructing *separate price indexes for land and for structures*. There is a multicollinearity problem between structure size and land plot size: large structures tend to be associated with large plots. This multicollinearity problem shows up in this section, where none of the straightforward methods suggested work. Thus in the next two sections, restrictions are imposed upon the hedonic regressions. In section 8, the price of constant quality structures is forced to be nondecreasing while in section 9, the price movements in constant quality structure prices are forced to follow the movements in an exogenous index of new dwelling construction costs. Both methods seem to work reasonably well but the results they generate are somewhat inconsistent.

A problem with many hedonic regression models is that historical results will generally change as new data become available. This problem is addressed by applying a *rolling window hedonic regression methodology* that is a generalization of the usual adjacent period time dummy hedonic regression methodology. This methodology is explained and illustrated in section 10.

Finally, in section 11, we show how the hedonic regression models for the *sales* of properties developed in sections 6 and 9 can be adapted to generate indexes for the *stock* of housing properties.

Section 12 offers some tentative conclusions.

2. Stratification Methods

A dwelling unit has a number of important *price determining characteristics*:

- The *land area* L of the property;
- The *floor space area* S of the structure; i.e., the size of the structure that sits on the land underneath and surrounding the structure;
- The *age* A of the structure, since this determines (on average) how much physical deterioration or depreciation the structure has experienced;
- The amount of *renovations* that have been undertaken for the structure;
- The *location* of the structure; i.e., its distance from amenities such as shopping centers, schools, restaurants and work place locations;
- The *type of structure*; i.e., single detached dwelling unit, row housing, low rise apartment or high rise apartment or condominium;
- The *type of construction* used to build the structure;
- Any *other special price determining characteristics* that are different from "average" dwelling units in the same general location such as swimming pools, air conditioning, elaborate landscaping, the height of the structure or views of oceans or rivers.

The data used in this study consist of observations on quarterly sales of detached houses for a small town (the population is around 60,000) in the Netherlands, town "A", for 14 quarters, starting in the first quarter of 2005 and ending in the second quarter of 2008. The variables used in this study can be described as follows:[3]

- p_n^t is the selling price of property n in quarter t in Euros where t = 1,...,14;
- L_n^t is the area of the plot for the sale of property n in quarter t in meters squared;
- S_n^t is the living space area of the structure for the sale of property n in quarter t in meters squared;
- A_n^t is the (approximate) age (in decades) of the structure on property n in quarter t.

The values of the fourth variable listed above are determined as follows. The original data were coded as follows: if the structure was built in 1960-1970, the observation was assigned the decade indicator variable BP = 5; 1971-1980, BP=6; 1981-1990, BP=7; 1991-2000, BP=8; 2001-2008, BP=9. The age variable A in this study was set equal to 9

[3] Houses that were older than 50 years at the time of sale were deleted from the data set. Two observations that had unusually low selling prices (36,000 and 40,000 Euros) were deleted as were 28 observations that had land areas greater than 1200 m^2. No other outliers were deleted from the sample.

– BP. For a recently built structure n in quarter t, $A_n^t = 0$. Thus the age variable gives the (approximate) age of the structure in decades.

It can be seen that not all of the price determining characteristics of the dwelling unit were used in the present study. In particular, the last five sets of price determining characteristics of the property listed above were neglected. Thus there is an implicit assumption that quarter to quarter changes in the amount of renovations that have been undertaken for the structures sold, the location of the structures, the type of structure, the type of construction used to build the structures and any other special price determining characteristics of the properties sold in the quarter did not change enough to be a significant determinant of the average price for the properties sold once changes in land size, structure size and the age of the structures were taken into account. To support this assumption, it should be noted that the hedonic regression models to be discussed later in the paper consistently explained 80-90% of the variation in the price data using just the three main explanatory variables: L, S and A.[4]

As mentioned above, there were 2289 observations on detached house sales for city "A" over the 14 quarters in the sample. Thus there was an average of 163.5 sales of detached dwelling units in each quarter. The overall sample mean selling price was 190,130 Euros, while the corresponding median price was 167,500 Euros. The average lot or plot size was 257.6 m^2 and the average size of structure was 127.2 m^2. The average age of the properties sold was approximately 18.5 years old.

The *stratification approach* to the construction of a house price index is conceptually very simple: for each important price explaining characteristic, divide up the sales into relatively homogeneous groups. Thus in the present case, sales were classified into 45 groups or cells consisting of 3 groupings for the land area L, 3 groupings for the structure area S and 5 groups for the age A (in decades) of the structure that was sold (3×3×5 = 45 separate cells). Once quarterly sales were classified into the 45 groupings of sales, the sales within each cell in each quarter were summed and then divided by the number of units sold in that cell in order to obtain unit value prices. These unit value prices were then combined with the number of units sold in each cell to form the usual p's and q's that can be inserted into a bilateral index number formula, like the Laspeyres (1871), Paasche (1874) and Fisher (1922) ideal formulae,[5] yielding a stratified index of house prices of each of these types.[6]

[4] The R^2 between the actual and predicted selling prices ranged from .83 to .89. The fact that it was not necessary to introduce more price determining characteristics for this particular data set can perhaps be explained by the nature of the location of the town of "A" on a flat, featureless plain and the relatively small size of the town; i.e., location was not a big price determining factor since all locations have basically the same access to amenities.

[5] The various international manuals on price measurement recommend this unit value approach to the construction of price indexes at the first stage of aggregation; see ILO, IMF, OECD, UNECE, Eurostat and World Bank (2004) and IMF, ILO, OECD, Eurostat, UNECE and the World Bank (2004) (2009). However, the unit value aggregation is supposed to take place over homogeneous items and this assumption may not be fulfilled in the present context, since there is a fair amount of variability in L, S and A within each cell. But since there is only a small number of observations in each cell for the data set under consideration, it

How should the size limits for the L and S groupings be chosen? One approach would be to divide the range of L and S by three and then create three equal size cells. However, this approach leads to a very large number of observations in the middle cells. Thus in the present study, size limits were chosen so that roughly 50% of the observations would fall into the middle sized categories and roughly 25% would fall into the small and large categories. For the land size variable L, the cutoff points chosen were 160 m^2 and 300 m^2, while for the structure size variable S, the cutoff points chosen were 110 m^2 and 140 m^2. Thus if $L < 160\ m^2$, then the observation fell into the small land size cell; if $160\ m^2 \le L < 300\ m^2$, then the observation fell into the medium land size cell and if $300\ m^2 \le L$, then the observation fell into the large land size cell. The resulting sample probabilities for falling into these three L cells over the 14 quarters were .24, .51 and .25 respectively. Similarly, if $S < 110\ m^2$, then the observation fell into the small structure size cell; if $110\ m^2 \le S < 140\ m^2$, then the observation fell into the medium structure size cell and if $140\ m^2 \le S$, then the observation fell into the large structure size cell. The resulting sample probabilities for falling into these three S cells over the 14 quarters were .21, .52 and .27 respectively.

The data that were used did not have an exact age for the structure; only the decade when the structure was built was recorded. Thus there was no possibility of choosing exact cutoff points for the age of the structure. For the first age group, A = 0 corresponds to a house that was built during the years 2001-2008; A = 1 for houses built during the years 1991-2000; A = 2 for houses build in 1981-1990, A = 3 for houses built in 1971-1980; and A = 4 for houses built in 1961-1970. The resulting sample probabilities for falling into these five cells over the 14 quarters were .15, .32, .21, .20 and .13 respectively. See Table 1 below for the sample joint probabilities of a house sale belonging to each of the 45 cells.

Table 1: Sample Probability of a Sale in Each Stratified Cell

		A = 0	A = 1	A = 2	A = 3	A = 4
L=small	S=small	0.00437	0.02665	0.01660	0.02053	0.02097
L=medium	S=small	0.00349	0.02840	0.01966	0.01092	0.03888
L=large	S=small	0.00087	0.00175	0.00044	0.00218	0.00612
L=small	S=medium	0.01223	0.05242	0.04281	0.02053	0.00699
L=medium	S=medium	0.03277	0.09262	0.08869	0.07907	0.02141
L=large	S=medium	0.00786	0.02315	0.01005	0.01442	0.01398
L=small	S=large	0.00306	0.00218	0.00175	0.00568	0.00000
L=medium	S=large	0.03145	0.03495	0.00786	0.02097	0.00306
L=large	S=large	0.04893	0.05461	0.02315	0.02490	0.01660

would be difficult to introduce more cells to improve homogeneity since this would lead to an increased number of empty cells and a lack of matching for the cells.

[6] However, since there are only 163 or so observations for each quarter and 45 cells to fill, it can be seen that each cell will have only an average of 3 or so observations in each quarter, and some cells were empty for some quarters. This problem will be addressed subsequently.

There are several points of interest to note about the above Table:

- There were no observations for houses built during the 1960s (the A = 4 class) that had a small lot (L = small) and a large structure (S = large), so this cell is entirely empty;
- There are many cells that are almost empty; in particular the probability of a sale of a large plot with a small house is very low as is the probability of a sale of a small plot with a large house;[7]
- The most *representative model* that is sold over the sample period corresponds to a medium sized lot, a medium sized structure and a house that was built in the 1990s (the A = 1 category). The sample probability of a house sale falling into this cell is 0.09262, which is the highest probability cell.

The average selling price of a house that falls into the medium L, medium S and A = 1 category is graphed in Figure 1 below along with the mean and median price of a sale in each quarter. These average prices have been converted into indexes that start at 1 for quarter 1, which is the first quarter of 2005. It should be noted that these three house price indexes are rather variable!

Some additional indexes are plotted in Figure 1, including a *fixed base matched model Fisher ideal index* and a *chained matched model Fisher ideal price index*. It is necessary to explain what a matched model index in this context means. If at least one house sold in each quarter for each of the 45 classes of transaction, then the ordinary *Laspeyres*, *Paasche* and *Fisher* price indexes, $P_L(s,t)$, $P_P(s,t)$ and $P_F(s,t)$, that compared the data in quarter s (in the denominator) to the data in quarter t (in the numerator) would be defined as follows:

$$(1)\ P_L(s,t) \equiv \textstyle\sum_{n=1}^{45} p_n^t q_n^s / \sum_{n=1}^{45} p_n^s q_n^s\ ;$$
$$(2)\ P_P(s,t) \equiv \textstyle\sum_{n=1}^{45} p_n^t q_n^t / \sum_{n=1}^{45} p_n^s q_n^t\ ;$$
$$(3)\ P_F(s,t) \equiv [P_L(s,t)P_P(s,t)]^{1/2}$$

where q_n^t is the *number of properties transacted* in quarter t in cell n and p_n^t is defined as the sum of the *values* for all properties transacted in quarter t in cell n divided by q_n^t and thus p_n^t is the *unit value price* for all properties transacted in cell n during quarter t for t = 1,...,14 and n = 1,...,45.

The above algebra is applicable to the case where there are transactions in all cells for the two quarters being compared. But for the present data set, on average only about 30 out of the 45 cell categories can be matched across any two quarters s and t. The above formulae (1)-(3) need to be modified to deal with this lack of matching problem. Thus when considering how to form an index number comparison between quarters s and t,

[7] Thus lot size and structure size are positively correlated with a correlation coefficient of .6459. Both L and S are fairly highly correlated with the selling price variable P: the correlation between P and L is .8234 and between P and S is .8100. These high correlations lead to some multicollinearity problems in the hedonic regression models to be considered later.

define *the set of cells n that have at least one transaction in each of quarters s and t* as the set S(s,t). Then the *matched model counterparts*, $P_{ML}(s,t)$, $P_{MP}(s,t)$ and $P_{MF}(s,t)$, to the indexes defined by (1), (2) and (3) are defined as follows:[8]

(4) $P_{ML}(s,t) \equiv \sum_{n\in S(s,t)} p_n^t q_n^s / \sum_{n\in S(s,t)} p_n^s q_n^s$;
(5) $P_{MP}(s,t) \equiv \sum_{n\in S(s,t)} p_n^t q_n^t / \sum_{n\in S(s,t)} p_n^s q_n^t$;
(6) $P_{MF}(s,t) \equiv [P_{ML}(s,t)P_{MP}(s,t)]^{1/2}$.

In Figure 1, the Fixed Base Fisher index is the matched model Fisher index defined by (6), where the base quarter s is kept fixed at quarter 1; i.e., the indexes $P_{MF}(1,1)$, $P_{MF}(1,2)$, ...,$P_{MF}(1,14)$ are calculated and labelled as the Fixed Base Fisher Index, P_{FFB}. The index that is labelled the Chained Fisher Index, P_{FCH}, is the index $P_{MF}(1,1)$, $P_{MF}(1,1)P_{MF}(1,2)$, $P_{MF}(1,1)P_{MF}(1,2)P_{MF}(2,3)$, ..., $P_{MF}(1,1)P_{MF}(1,2)P_{MF}(2,3)P_{MF}(3,4)$ ... $P_{MF}(13,14)$. Note that the Fixed Base and Chained Fisher (matched model) indexes are quite close to each other and are *much smoother* than the corresponding Mean, Median and Representative Model indexes.[9] The data for the five series defined thus far are listed in Table 2 below along with two additional series, P_{IFCH} and P_{IFFB}, which will be defined shortly. The seven series are plotted in Figure 1 below.

Table 2: Matched Model Fisher Chained and Fixed Base Indexes, With and Without Price Imputation , Mean, Median and Representative Model House Price Indexes

Quarter	P_{FCH}	P_{IFCH}	P_{FFB}	P_{IFFB}	P_{Mean}	P_{Median}	$P_{Represent}$
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.02396	1.02518	1.02396	1.02518	1.02003	1.05806	1.04556
3	1.07840	1.07827	1.06815	1.07354	1.04693	1.02258	1.03119
4	1.04081	1.03781	1.04899	1.04754	1.05067	1.03242	1.04083
5	1.04083	1.03763	1.04444	1.04352	1.04878	1.04839	1.04564
6	1.05754	1.05604	1.06676	1.06329	1.13679	1.17581	1.09792
7	1.07340	1.07536	1.07310	1.07686	1.06490	1.06935	1.01259
8	1.06706	1.06548	1.07684	1.07777	1.07056	1.10000	1.10481
9	1.08950	1.08932	1.06828	1.07335	1.07685	1.05806	1.03887
10	1.11476	1.11388	1.11891	1.11724	1.16612	1.16048	1.07922
11	1.12471	1.12445	1.12196	1.12139	1.08952	1.06290	1.07217
12	1.10483	1.10532	1.11321	1.11054	1.09792	1.10323	1.03870
13	1.10450	1.10511	1.11074	1.10583	1.10824	1.12903	1.12684

[8] A justification for this approach to dealing with a lack of matching in the context of bilateral index number theory can be found in the discussion by Diewert (1980; 498-501) on the related problem of dealing with new and disappearing goods. Other approaches are also possible. For approaches based on imputation methods, see Alterman, Diewert and Feenstra (1999) (and the discussion below) and for approaches that are based on maximum matching over all pairs of periods, see Ivancic, Diewert and Fox (2011) and Haan and van der Grient (2011).

[9] The means (and standard deviations) of the five series mentioned thus far are as follows: P_{FCH} = 1.0737 (0.0375), P_{FFB} = 1.0737 (0.0370), P_{Mean} = 1.0785 (0.0454), P_{Median} = 1.0785 (0.0510), and $P_{Represent}$ = 1.0586 (0.0366). Thus the representative model price index has a smaller variance than the two matched model Fisher indexes but it has a substantial bias relative to these two Fisher indexes: the representative model price index is well below the Fisher indexes for most of the sample period.

14	1.11189	1.11224	1.10577	1.10654	1.12160	1.10323	1.08587

Figure 1: Fisher Matched Model With and Without Imputed Prices and Various Summary Statistic Indexes

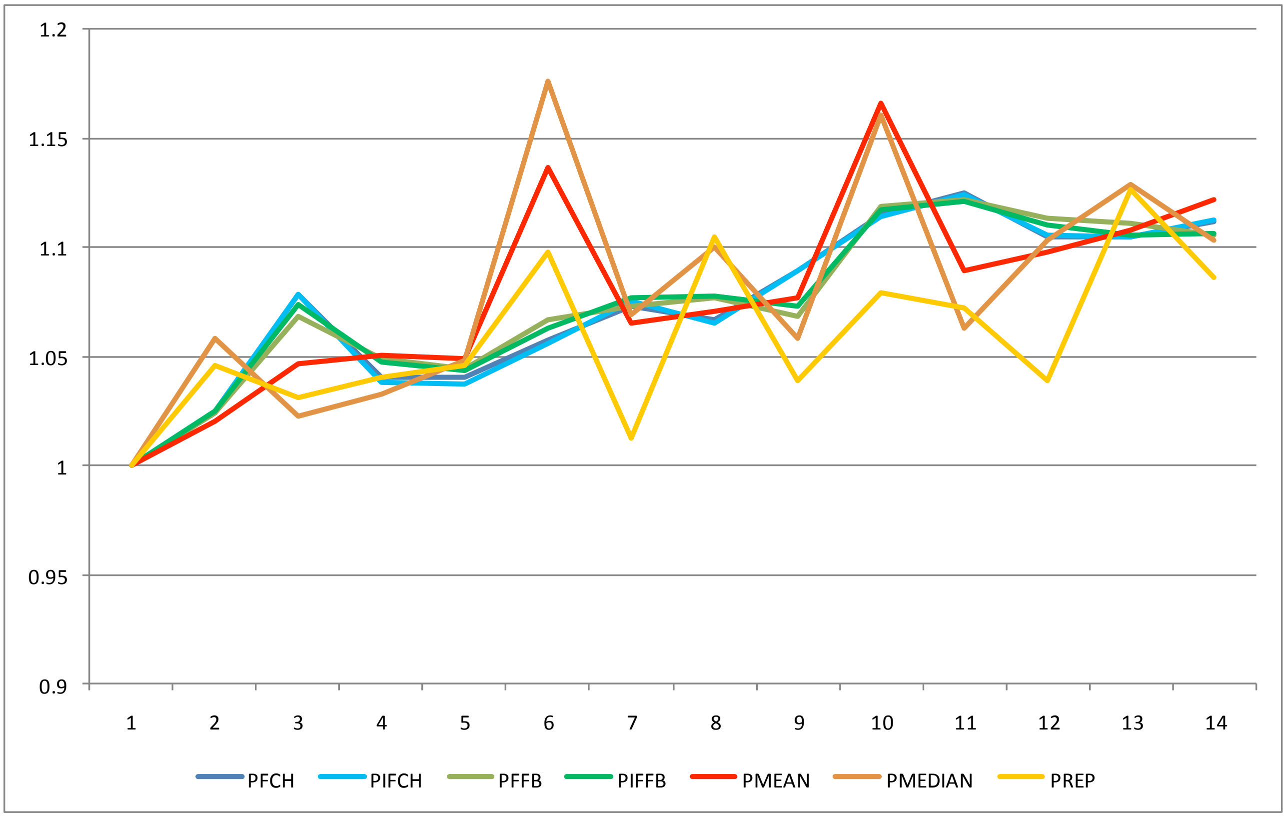


Table 2 and Figure 1 show two other series: P_{IFCH} and P_{IFFB}. In order to improve the degree of matching between any two periods, these two series make use of *imputed prices* for the cells that have no observations in any given quarter. In section 6, a simple hedonic regression model is estimated for each period; see equations (16) below. Let α^{t*}, β^{t*}, γ^{t*} and δ^{t*} be the *estimated coefficients for the quarter t regression* for t = 1,…,14. For each of the 45 cells in our stratification structure, define the entire sample *average amounts of land* L and *structures* S in cell i,j,k as $L_{i,j,k}$ and $S_{i,j,k}$ for i = 1,2,3;[10] j = 1,2,3; k = 1,…,5. Also define $A_{i,j,k} \equiv k-1$ as the *value of the age variable* in cell i,j,k for i = 1,2,3; j = 1,2,3; k= 1,…,5. Using these definitions, a set of 45 *imputed prices* can be defined for each cell i,j,k in the stratification scheme and each quarter t as follows:

$$(7)\ p_{i,j,k}{}^{t} \equiv \alpha^{t*} + \beta^{t*}L_{i,j,k} + \gamma^{t*}(1 - \delta^{t*}A_{i,j,k})S_{i,j,k}\ ; \quad i = 1,2,3;\ j = 1,2,3;\ k = 1,\ldots,5;\ t = 1,\ldots,14.$$

[10] Cell i = 1 (L is small), j = 3 (S is large) and k = 5 (A is old; i.e., A = 4) is empty since no houses of this type sold over the 14 quarters in our sample. We arbitrarily set $L_{1,3,5} \equiv L_{1,3,4}$ and $S_{1,3,5} \equiv S_{1,3,4}$. These definitions will not affect the subsequent Laspeyres, Paasche and Fisher indexes.

Now recall the definitions for the Laspeyres, Paasche and Fisher price indexes, $P_L(s,t)$, $P_P(s,t)$ and $P_F(s,t)$, that compared the data in quarter s (in the denominator) to the data in quarter t (in the numerator), (1), (2) and (3) above. If in quarter t, cell n in these formulae turned out to be empty, then the cell n unit value p_n^t and the corresponding quantity transacted q_n^t were defined to be zeros. We leave the quantity variables unchanged but if a unit value price p_n^t is zero, redefine it to be the corresponding imputed price $p_{i,j,k}^t$ defined by (7) that corresponds to the cell n. With these changes, all prices are now positive and formulae (1)-(3) can be used without modification for matching in order to construct Laspeyres, Paasche and Fisher indexes between periods s and t. Thus P_{IFCH} and P_{IFFB} which appear in Table 2 and Figure 1 are the resulting *chained and fixed base stratified sample Fisher house price indexes that use imputed prices* for missing cell prices. It can be seen that the new chained Fisher index that uses imputed prices P_{IFCH} is extremely close to its counterpart P_{FCH} that uses only matched observations and the new fixed base Fisher index that uses imputed prices P_{IFFB} is very close to its counterpart P_{FFB}.[11] For this particular data set, the use of imputed prices to improve the degree of period to period matching of prices did not make much of a difference.

The four matched model Fisher indexes must be regarded as being more accurate than the other indexes, which use only a limited amount of the available price and quantity information. Any one of the Fisher indexes could be used as a headline index of house price inflation. Since all of the Fisher indexes trend fairly smoothly, the two chained Fisher indexes should be preferred over the two fixed base Fisher indexes, following the advice in Hill (1988) (1993) and in the CPI Manual; see the ILO, IMF, OECD, UNECE, Eurostat and World Bank (2004). Note also that there is no need to use Laspeyres or Paasche indexes in this situation since real estate data on sales of houses contains both value and quantity information. Under these conditions, Fisher indexes are preferred by the above sources over the Laspeyres and Paasche indexes (which do not use all of the available price and quantity information for the two periods being compared).

Since there is a considerable amount of heterogeneity in each cell of the stratification scheme, there is the strong possibility of some *unit value bias*[12] in the matched model Fisher indexes. However, if a finer cell classification were used, the amount of matching would drop dramatically. Already, with the present classification, only about 2/3 of the cells could be matched across any two quarters. Thus there is a tradeoff between having too few cells with the possibility of unit value bias and having a finer cell classification

[11] The correlation coefficient between P_{FCH} and P_{IFCH} is .99929 and the correlation between P_{FFB} and P_{IFFB} is .99672. The correlation coefficients between P_{FCH} and P_{FFB} and P_{IFFB} are .97292 and .98445; the correlation coefficients between P_{IFCH} and P_{FFB} and P_{IFFB} are .96964 and .98210. Thus the use of imputed prices to improve the degree of matching has narrowed the differences between the chained and fixed base Fisher indexes. The sample mean (and standard deviation) of P_{IFCH} is 1.0733 (0.0379) and for P_{IFFB} is 1.0745 (0.0363).

[12] See Balk (1998) (2008; 72-74), Silver (2009a) (2009b) (2010) and Diewert and von der Lippe (2010) for discussions of unit value bias.

scheme but with a much smaller degree of matching of the data within cells across the two time periods being compared.[13]

Looking at Table 2 and Figure 1, it can be seen that the two chained Fisher indexes considered above show drops in house prices in the fourth quarter of 2005, 2006 and 2007. Thus there is the possibility that house prices drop for seasonal reasons in the fourth quarter of each year. In order to deal with this possibility, a rolling year matched model Fisher index is constructed in the following section.

3. Rolling Year Indexes and Seasonality

Assuming that each commodity in each season of the year is a separate "annual" commodity is the simplest and theoretically most satisfactory method for dealing with seasonal commodities when the goal is to construct *annual* price and quantity indexes. This idea can be traced back to Mudgett in the consumer price context and to Stone in the producer price context:

"The basic index is a yearly index and as a price or quantity index is of the same sort as those about which books and pamphlets have been written in quantity over the years." Bruce D. Mudgett (1955; 97).

"The existence of a regular seasonal pattern in prices which more or less repeats itself year after year suggests very strongly that the varieties of a commodity available at different seasons cannot be transformed into one another without cost and that, accordingly, in all cases where seasonal variations in price are significant, the varieties available at different times of the year should be treated, in principle, as separate commodities." Richard Stone (1956; 74-75).

Diewert (1983) generalized the Mudgett-Stone annual framework to allow for *rolling year comparisons* for 12 consecutive months of data with a base year of 12 months of data or for comparisons of four consecutive quarters of data with a base year of 4 consecutive quarters of data; i.e., the basic idea is to compare the current rolling year of price and quantity data to the corresponding data of a base year where the data pertaining to each season is compared.[14] Thus in the present context, we have in principle,[15] price and quantity data for 45 classes of housing commodities in each quarter. If the sale of a house in each season is treated as a separate good, then there are 180 annual commodities.

For the first index number value, the four quarters of price and quantity data on sales of detached dwellings in the town of "A" (180 series) are compared with the same data

[13] Diewert and von der Lippe (2010) show that with finer and finer stratification schemes, eventually there is a complete lack of matching and index numbers based on highly stratified unit values become meaningless.

[14] For additional examples of this rolling year approach, see the chapters on seasonality in ILO, IMF, OECD, UNECE, Eurostat and World Bank (2004), the IMF, ILO, OECD, Eurostat, UNECE and the World Bank (2004) and Diewert (1998). In order to theoretically justify the rolling year indexes from the viewpoint of the economic approach to index number theory, some restrictions on preferences are required. The details of these assumptions can be found in Diewert (1999; 56-61). It should be noted that weather and the lack of fixity of Easter can cause "seasons" to vary and a breakdown in the approach; see Diewert, Finkel and Artsev (2009). However, with quarterly data, these limitations of the rolling year index are less important.

[15] In practice, as we have seen in the previous section, many of the cells are empty in each period.

using the Fisher ideal formula. Naturally, the resulting index is equal to 1. For the next index number value, the data for the first quarter of 2005 are dropped and the data pertaining to the first quarter of 2006 are appended to the data for quarters 2-4 of 2005. The resulting Fisher index is the second entry in the RY Matched Model series that is illustrated in Figure 2 below. However, as was the case with the chained and fixed base Fisher indexes that appeared in Figure 1 above, not all cells could be matched using the rolling year methodology; i.e., some cells were empty in the first quarter of 2006 which corresponded to cells in the first quarter of 2005 which were not empty and vice versa. Thus when constructing the rolling year index P_{RY} plotted in Figure 2, the comparison between the rolling year and the data pertaining to 2005 was restricted to the set of cells which were non empty in both years; i.e., the Fisher rolling year indexes plotted in Figure 2 are matched model indexes. Unmatched models are omitted from the index number comparison.[16] The results can be observed in Figure 2. Note that there is a definite downturn at the end of the sample period but that the downturns which showed up in Figure 1 for quarters 4 and 8 can be interpreted as seasonal downturns; i.e., the rolling year indexes in Figure 2 did not turn down until the end of the sample period. Note also that the index value for observation 5 compares the data for calendar year 2006 to the corresponding data for calendar year 2005 and the index value for observation 9 compares the data for calendar year 2007 to the corresponding data for calendar year 2005; i.e., these index values correspond to Mudgett-Stone annual indexes.

It is a fairly labour intensive job to construct the rolling year matched model Fisher indexes since the cells that are matched over any two periods vary with the periods. A short cut method for seasonally adjusting a series such as the matched model chained Fisher index P_{FCH} and the fixed base Fisher index P_{FFB} listed in Table 2 in the previous section is to simply take a 4 quarter *moving average* of these series. The resulting rolling year series, P_{FCHMA} and P_{FFBMA}, can be compared with the rolling year Mudgett-Stone-Diewert series P_{RY}; see Figure 2 below. The data that corresponds to Figure 2 are listed in Table 3 below.

Figure 2: Rolling Year Fixed Base Fisher P_{FFBRY}, Fisher Chained Moving Average P_{FCHMA} and Fisher Fixed Base Moving Average P_{FFBMA} House Price Indexes

[16] There are 11 rolling year comparisons that can be made with the data for 14 quarters that are available. The number of unmatched or empty cells for rolling years 2, 3, ..., 11 are as follows: 50, 52, 55, 59, 60, 61, 65, 65, 66, 67. The relatively low number of unmatched or empty cells for rolling years 2, 3 and 4 is due to the fact that for rolling year 2, ¾ of the data are matched, for rolling year 3, ½ of the data are matched and for rolling year 4, ¼ of the data are matched.

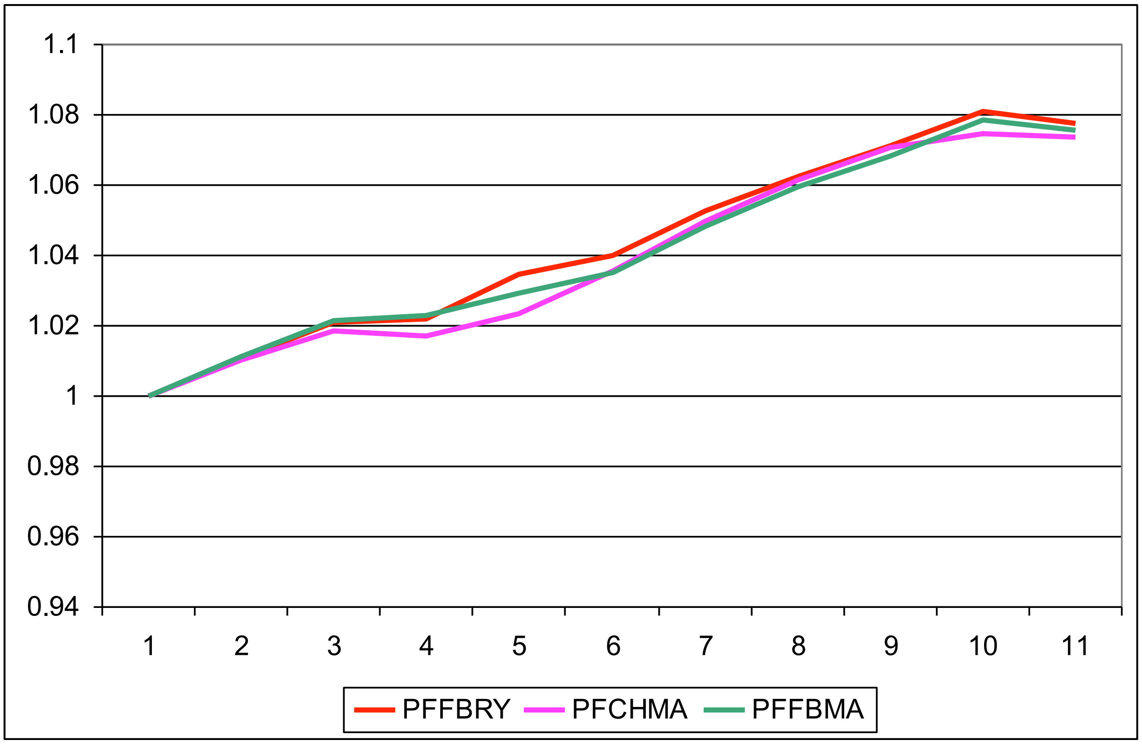


Table 3: Rolling Year Fixed Base Fisher P_{FFBRY}, Fisher Chained Moving Average P_{FCHMA} and Fisher Fixed Base Moving Average P_{FFBMA} House Price Indexes

Rolling Year	P_{FFBRY}	P_{FCHMA}	P_{FFBMA}
1	1.00000	1.00000	1.00000
2	1.01078	1.01021	1.01111
3	1.02111	1.01841	1.02156
4	1.02185	1.01725	1.02272
5	1.03453	1.02355	1.02936
6	1.04008	1.03572	1.03532
7	1.05287	1.04969	1.04805
8	1.06245	1.06159	1.05948
9	1.07135	1.07066	1.06815
10	1.08092	1.07441	1.07877
11	1.07774	1.07371	1.07556

It can be seen that a simple moving average of the chained Fisher and fixed base quarter to quarter indexes, P_{FCH} and P_{FFB}, listed in Table 2 of the previous section approximates the theoretically preferred rolling year fixed base Fisher index P_{FFBRY} fairly well. However, there are differences of up to 1% between the preferred rolling year index and the moving average index. Recall that the fixed base Fisher index constructed in the previous section compared the data of quarters 1 to 14 with the corresponding data of quarter 1. Thus the observations for, say, quarters 2 and 1, 3 and 1, and 4 and 1 are not as likely to be as comparable as the rolling year indexes where data in any one quarter is always lined up with the data in the corresponding quarter of the base year. A similar argument applies to the moving average index P_{FCHMA}; the comparisons that go into the links in this index are from quarter to quarter and they are unlikely to be as accurate as comparisons across the years for the same quarter.[17]

We turn now to methods for constructing house price indexes that are based on hedonic regression techniques.

4. Time Dummy Hedonic Regression Models using the Logarithm of Price as the Dependent Variable

The most popular hedonic regression models regress the log of the price of the good on either a linear function of the characteristics or on the logs of the characteristics along with time dummy variables.[18] We will consider each of these models in turn.

The Log Linear Time Dummy Hedonic Regression Model:

In quarter t, there were N(t) sales of detached houses in the town of "A" where p_n^t is the selling price of house n sold during quarter t. We have information on three characteristics of house n sold in period t: L_n^t is the area of the plot in square meters (m^2); S_n^t is the floor space area of the structure in m^2 and A_n^t is age in decades of house n in period t. The *Log Linear time dummy hedonic regression model* is defined by the following system of regression equations:[19]

$$(8)\ \ln p_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \tau^t + \varepsilon_n^t\ ; \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0$$

where τ^t is a quarter t shift parameter which shifts the hedonic surface upwards or downwards as compared to the quarter 1 surface.[20]

Note that if we exponentiate both sides of (8) and neglect the error term, then the house price p_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

[17] The stronger is the seasonality, the stronger will be this argument in favour of the accuracy of the rolling year index. The strength of this argument can be seen if all house price sales in any given cell turn out to be *strongly seasonal*; i.e., the sales for each cell occur in only one quarter in each year. Quarter to quarter comparisons are obviously impossible in this situation but rolling year indexes will be perfectly well defined.

[18] This methodology was developed by Court (1939; 109-111) as his hedonic suggestion number two but there were earlier contributions that were not noticed by the profession until recently.

[19] For all the models estimated in this paper, it is assumed that the error terms ε_n^t are independently distributed normal variables with mean 0 and constant variance and maximum likelihood estimation is used in order to estimate the unknown parameters in each regression model. The nonlinear option in Shazam was used for the actual estimation.

[20] The 15 parameters $\alpha, \tau^1,\ldots,\tau^{14}$ correspond to variables that are exactly collinear in the regression (8) and thus the restriction $\tau^1 = 0$ is imposed in order to identify the remaining parameters.

house price index using this model; see Figure 3 and Table 4 below for the resulting index which is labelled as P_{H1} (hedonic house price index 1). The R^2 for this model was .8420 which is quite satisfactory for a hedonic regression model with only three characteristics. For later comparison purposes, we note that the log likelihood was 1407.6.

A problem with this model is that the underlying price formation model seems implausible: S and L interact multiplicatively in order to determine the overall house price whereas it seems likely that lot size L and house size S interact in an approximately additive fashion to determine the overall house price.

Another problem with the regression model (8) is that age is entered in an additive fashion. The problem with this is that we would expect age to interact directly with the structures variable S as a (net) depreciation variable (and not interact directly with the land variable, which does not depreciate). In the following model, we make this direct interaction adjustment to (8).

The Log Linear Time Dummy Hedonic Regression Model with Quality Adjustment of Structures for Age

In this model, we argue that age A interacts with the quantity of structures S in a multiplicative manner; i.e., an appropriate explanatory variable for the selling price of a house is $\gamma(1-\delta)^A S$ (geometric depreciation where δ is the decade geometric depreciation rate) or $\gamma(1-\delta A)S$ (straight line depreciation where δ is the decade straight line depreciation rate) instead of the additive specification $\gamma S + \delta A$. In what follows, the straight line variant of this class of models is estimated[21]; i.e., the *Log Linear time dummy hedonic regression model with quality adjusted structures* is the following regression model:

$$(9)\ \ln p_n^t = \alpha + \beta L_n^t + \gamma(1 - \delta A_n^t)S_n^t + \tau^t + \varepsilon_n^t\,; \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0.$$

The above regression model was run using the 14 quarters of sales data for the town of "A". Note that only one common straight line depreciation rate δ is estimated. The estimated decade (net) depreciation rate[22] was $\delta^* = 11.94\%$ (or around 1.2% per year), which is very reasonable. As was the case with the previous model, 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; see Figure 3 and Table 4 below (see P_{H2}) for the resulting index. The R^2 for this model was .8345, a bit lower than the previous model and the log likelihood was 1354.9, which is quite a drop from the previous log likelihood of 1407.6. Thus it appears that the imposition of more theory (with respect to the treatment of the age of the house) has led to a drop in the empirical fit of the model.

---

[21] This regression is essentially linear in the unknown parameters and hence it is very easy to estimate.

[22] It is a net depreciation rate because we have no information on renovation expenditures so δ serves as a net depreciation rate; i.e., it is equal to gross wear and tear depreciation of the house less average real expenditures on renovations and repairs.

However, it is likely that this model and the previous one are misspecified[23]: they both multiply together land area times structure area in order to determine the price of the house and it is likely that an additive interaction between L and S is more appropriate than a multiplicative one.

Note that once the depreciation rate has been estimated (denote the estimated rate by δ^*), then *quality adjusted structures* (adjusted for the aging of the structure) for each house n in each quarter t can be defined as follows:

$$(10)\ S_n^{t*} \equiv (1 - \delta^* A_n^t) S_n^t\,; \qquad t = 1,...,14;\ n = 1,...,N(t).$$

The Log Log Time Dummy Hedonic Regression Model with Quality Adjustment of Structures for Age

In this model, we will work with quality adjusted (for age) structures, $(1-\delta A)S$, rather than the unadjusted structures area, S. The Log Log model is similar to the previous Log Linear model, except that now, instead of using L and $(1-\delta A)S$ as explanatory variables in the regression model, we use the logarithms of the land and quality adjusted structures areas as independent variables. Thus the *Log Log time dummy hedonic regression model with quality adjusted structures* is the following regression model:

$$(11)\ \ln p_n^t = \alpha + \beta \ln L_n^t + \gamma \ln[(1 - \delta A_n^t) S_n^t] + \tau^t + \varepsilon_n^t; \qquad t = 1,...,14;\ n = 1,...,N(t);\ \tau^1 \equiv 0.$$

Using the data for "A", the estimated decade (net) depreciation rate[24] was $\delta^* = 0.1050$ (standard error 0.00374), which is a reasonable decade net depreciation rate. Note that if we exponentiate both sides of (11) and neglect the error term, the house price p_n^t would equal $e^\alpha [L_n^t]^\beta [S_n^{t*}]^\gamma [\exp\tau^t]$ where S_n^{t*} is defined as quality adjusted structures, $(1-\delta A_n^t)S_n^t$. 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 again can serve as the chain link in a price index; see Figure 3 and Table 4 below (see P_{H3}) for the resulting index. The R^2 for this model was .8599, which is a big increase over the previous two models and the log likelihood was 1545.4, a huge increase over the log likelihoods for the previous two models (1407.6 and 1354.9).

---

[23] If the variation in the independent variables is relatively small, the difference in indexes generated by the various hedonic regression models considered in this section and the following sections is likely to be small since virtually all of the models considered can offer roughly a linear approximation to the "truth". But when the variation in the independent variables is large (as it is in the present housing context), then the choice of functional form can have a very substantial effect. Thus a priori reasoning should be applied to both the choice of independent variables in the regression as well as to the choice of functional form. For additional discussion on functional form issues, see Diewert (2003a).

[24] It is a net depreciation rate because we have no information on renovation expenditures so δ is equal to average gross wear and tear depreciation of the house less average real expenditures on renovations and repairs.

It turns out that this hedonic regression model is a variant of McMillen's (2003) *consumer oriented approach to hedonic housing models*. It is worthwhile outlining his theoretical framework.[25]

A very simple way to justify a hedonic regression model from a consumer perspective is to postulate that households 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 "model" with characteristics z_1, z_2 yielding the scalar welfare measure, $f(z_1,z_2)$. Thus households 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 $f(z_1^1,z_2^1) > f(z_1^2,z_2^2)$.[26] Thus having more of every characteristic is always preferred by households. The next assumption that we make is that in period t, there is a positive *generic price for all models*, 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]

$$(12)\ W^t(z_{1n}^t,z_{2n}^t) = \rho^t\, f(z_{1n}^t,z_{2n}^t) = p_n^t.$$

The above willingness to pay for a house is set equal to the selling price of the house, p_n^t. Now all that is necessary is to specify the z characteristics and pick a functional form for the (cardinal) utility function f. In order to relate (12) to (11), let $z_{1n}^t \equiv L_n^t$ and $z_{2n}^t \equiv [(1 - \delta A_n^t)S_n^t$ and let $f(z_1,z_2)$ be the following *Cobb-Douglas utility function*:

$$(13)\ f(z_1,z_2) \equiv e^{\alpha}z_1^{\beta}z_2^{\gamma}\ ;\ \beta > 0\ ;\ \gamma > 0.$$

Now define $\rho^t \equiv \exp\tau^t$ for t = 1,...,14 and it can be seen that with these definitions, the hedonic regression model defined by (12) is equivalent to the model defined by (11), neglecting the error terms.

If β and γ sum to one, then the consumer's characteristics utility function exhibits *constant returns to scale*. Thus if z_1 and z_2 are multiplied by the positive scalar λ, then the consumer's initial utility $f(z_1,z_2)$ is also multiplied by λ; i.e., we have $f(\lambda z_1,\lambda z_2) = \lambda f(z_1,z_2)$ for all $\lambda > 0$. For the data pertaining to the town of "A", we obtained the following estimates for β and γ (standard errors in brackets): $\beta^* = 0.4196$ (0.00748) and $\gamma^* = 0.5321$ (0.0157). Thus the sum of β^* and γ^* was 0.9517, which is reasonably close to one.

Although this model performs the best of the simple hedonic regression models considered thus far, it has the unsatisfactory feature that the quantity of land and quality

[25] This exposition follows that of Diewert, Haan and Hendriks (2010).

[26] It is natural to impose some regularity conditions on the characteristics aggregator function f such as 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 $f(0,0) = 0$.

[27] For more elaborate justifications for household based hedonic regression models, see Muellbauer (1974) and Diewert (2003a).

adjusted structures determine the price of a house in a *multiplicative manner* when it is more likely that house prices are determined by a weighted *sum* of their land and quality adjusted structures amounts. Thus in the following section, an additive time dummy hedonic regression model will be estimated and the expectation is that this model will fit the data better.

The three house price series generated by the three time dummy hedonic regressions described in this section where the logarithm of the selling price is used as the dependent variable, P_{H1}, P_{H2} and P_{H3}, are plotted in Figure 3 below along with the stratified sample matched model chained Fisher house price index described in section 2 above, P_{FCH}. These four house price series are listed in Table 4 below.

Figure 3: Three Time Dummy Hedonic Regression Based House Price Indexes P_{H1}, P_{H2} and P_{H3} and the Stratified Sample Matched Model Chained Fisher House Price Index P_{FCH}

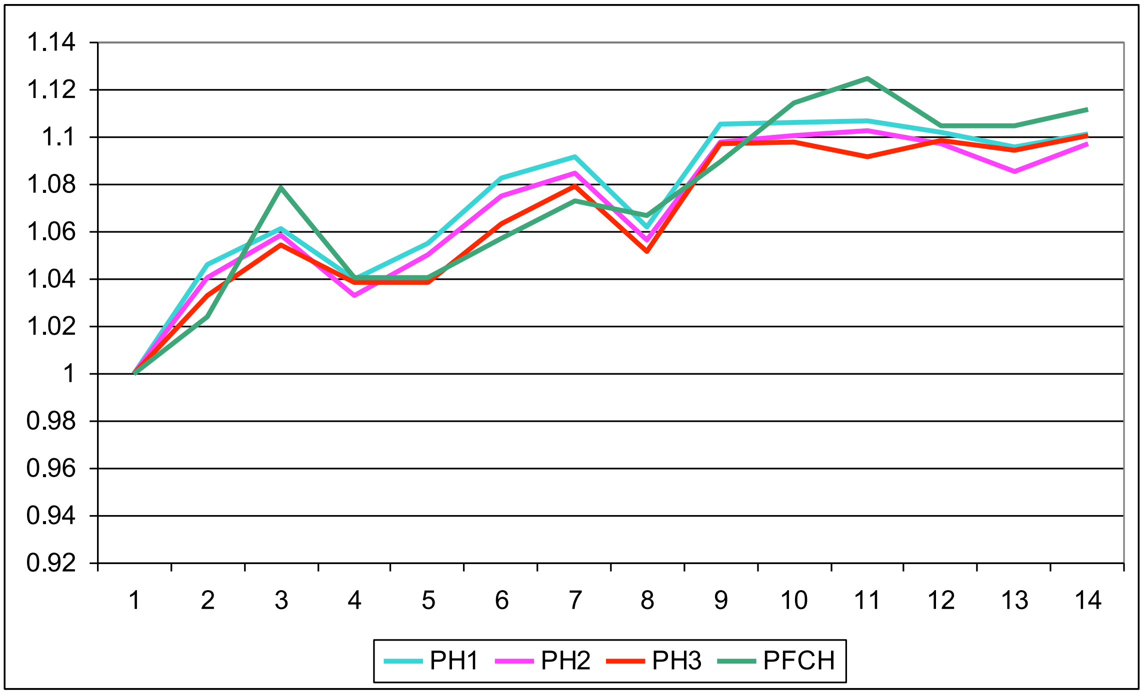


Table 4: Time Dummy House Price Indexes Using Hedonic Regressions with the Logarithm of Price as the Dependent Variable P_{H1}, P_{H2} and P_{H3} and the Stratified Sample Matched Model Chained Fisher Index P_{FCH}

Quarter	P_{H1}	P_{H2}	P_{H3}	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.04609	1.04059	1.03314	1.02396
3	1.06168	1.05888	1.05482	1.07840
4	1.04007	1.03287	1.03876	1.04081

5	1.05484	1.05032	1.03848	1.04083
6	1.08290	1.07532	1.06369	1.05754
7	1.09142	1.08502	1.07957	1.07340
8	1.06237	1.05655	1.05181	1.06706
9	1.10572	1.09799	1.09736	1.08950
10	1.10590	1.10071	1.09786	1.11476
11	1.10722	1.10244	1.09167	1.12471
12	1.10177	1.09747	1.09859	1.10483
13	1.09605	1.08568	1.09482	1.10450
14	1.10166	1.09694	1.10057	1.11189

It can be seen that all four indexes capture the same trend but there can be differences of over 2 percent between the various indexes for some quarters. Note that all of the indexes move in the same direction from quarter to quarter with decreases in quarters 4, 8, 12 and 13 except that P_{H3} (the index that corresponds to the Log Log model) increases in quarter 12.

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

The Linear Time Dummy Hedonic Regression Model

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.[28] 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:

$$(14)\quad p_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \tau^t + \varepsilon_n^t\,; \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0.$$

The above linear regression model was run using the data for the town of "A". The R^2 for this model was .8687, much higher than those obtained in our previous regressions and the log likelihood was −10790.4 (which cannot be compared to the previous log likelihoods since the dependent variable has changed from the logarithm of price to just price). Using model (14) to form an overall house price index is a bit more difficult than using the time dummy regression models in the previous section. In the previous section, holding characteristics constant and neglecting error terms, the *relative price* for the same model over any two time periods turned out to be constant, leading to an unambiguous overall index. In the present section, holding characteristics constant and neglecting error terms, the *difference in price* for the same model turns out to be constant, but the relative prices for different models will not in general be constant. Thus an overall index will be constructed which uses the prices generated by the estimated parameters in (14) and evaluated at the sample average amounts of L, S and the average age of a house A.[29] The

[28] Diewert (2007) and Diewert, Haan and Hendriks (2010) develop this line of thought in more detail.
[29] The sample average amounts of L and S were 257.6 m^2 and 127.2 m^2 respectively and the average age of the detached dwellings sold over the sample period was 1.85 decades.

resulting quarterly house prices for this "average" model were converted into an index, P_{H4}, which is listed in Table 5 below and charted in Figure 4.

The hedonic regression model defined by (14) is perhaps the simplest possible one but it is a bit too simple since it neglects the fact that the interaction of age with the selling price of the property takes place via a multiplicative interaction with the structures variable and not via a general additive factor. Thus in the following section, we will 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.

The Linear Time Dummy Hedonic Regression Model with Quality Adjusted Structures

The *linear time dummy hedonic regression model with quality adjusted structures* is the following regression model:

$$(15)\ p_n^t = \alpha + \beta L_n^t + \gamma(1 - \delta A_n^t)S_n^t + \tau^t + \varepsilon_n^t\ ; \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0.$$

This is the most plausible hedonic regression model so far. It works with quality adjusted (for age) structures S^* equal to $(1-\delta A)S$ instead of having A and S as completely independent variables that enter into the regression in a linear fashion.

The results for this hedonic regression model were a clear improvement over the results of the previous model, (14). The log likelihood increased by 92 to −10697.8 and the R^2 increased to .8789 from the previous .8687. The estimated decade depreciation rate was $\delta^* = 0.1119$ (0.00418), which is reasonable as usual. This linear regression model has the same property as the previous model: house price *differences* are constant over time for all constant characteristic models but house price *ratios* are not constant. Thus as in the previous model, an overall index will be constructed that uses the prices generated by the estimated parameters in (15) 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_{H5}, which is listed in Table 5 below and charted in Figure 4. For comparison purposes, P_{H3} (the time dummy Log Log model index) and P_{FCH} (the stratified sample chained matched model Fisher index) will be charted along with P_{H4} and P_{H5}. Our preferred indexes are P_{FCH} and P_{H5}.

Figure 4: Two Time Dummy House Price Indexes Using Hedonic Regressions with Price as the Dependent Variable, P_{H4} and P_{H5}, the Log Log Time Dummy index P_{H3} and the Stratified Sample Matched Model Chained Fisher Index P_{FCH}

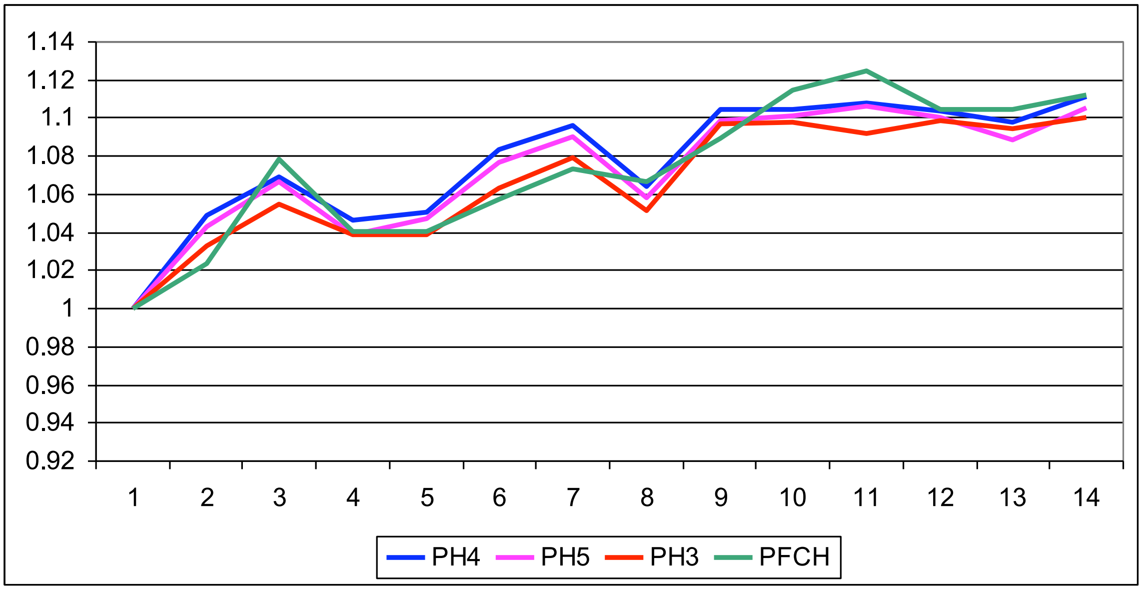


Table 5: Two Time Dummy House Price Indexes Using Hedonic Regressions with Price as the Dependent Variable, P_{H4} and P_{H2}, the Log Log Time Dummy index P_{H3} and the Stratified Sample Matched Model Chained Fisher Index P_{FCH}

Quarter	P_{H4}	P_{H5}	P_{H3}	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.04864	1.04313	1.03314	1.02396
3	1.06929	1.06667	1.05482	1.07840
4	1.04664	1.03855	1.03876	1.04081
5	1.05077	1.04706	1.03848	1.04083
6	1.08360	1.07661	1.06369	1.05754
7	1.09593	1.09068	1.07957	1.07340
8	1.06379	1.05864	1.05181	1.06706
9	1.10496	1.09861	1.09736	1.08950
10	1.10450	1.10107	1.09786	1.11476
11	1.10788	1.10588	1.09167	1.12471
12	1.10403	1.10044	1.09859	1.10483
13	1.09805	1.08864	1.09482	1.10450
14	1.11150	1.10572	1.10057	1.11189

It can be seen that again, all four indexes capture the same trend but there can be differences of over 2 percent between the various indexes for some quarters. Note that all of the indexes move in the same direction from quarter to quarter with decreases in quarters 4, 8, 12 and 13, except that P_{H3} increases in quarter 12.

A major problem with the hedonic time dummy regression models considered thus far is that the prices of land and quality adjusted structures are not allowed to change in an unrestricted manner from period to period. The class of hedonic regression models to be studied in the following section does not suffer from this problem.

6. Hedonic Imputation Regression Models

The theory of *hedonic imputation indexes* works as follows[30]: for each period, run a linear regression of the following form:

$$(16)\ p_n^{\,t} = \alpha^t + \beta^t L_n^{\,t} + \gamma^t(1 - \delta^t A_n^{\,t})S_n^{\,t} + \varepsilon_n^{\,t}\ ; \qquad t = 1,...,14;\ n = 1,...,N(t).$$

Note that there are only 4 parameters to be estimated for each quarter: α^t, β^t, γ^t and δ^t for t = 1,...,14.[31] Note also that (15) is similar in form to the model defined by equations (14), but with some significant differences:

- Only one depreciation parameter is estimated in the model defined by (15) whereas in the present model, there are 14 depreciation parameters, one for each quarter.
- In model (15), there was only one α, β, γ and δ parameter whereas in (16), there are 14 α^t, 14 β^t, 14 γ^t and 14 δ^t parameters to be estimated. On the other hand, model (14) had an additional 13 time shifting parameters (the τ^t) that required estimation.

Thus the hedonic imputation model involves the estimation of 56 parameters whereas the time dummy model required the estimation of only 17 parameters. Hence it is likely that the hedonic imputation model will fit the data much better.

As usual, in the housing context, we almost never have matched models across periods (there are always depreciation and renovation activities that make a house in the exact same location not quite comparable over time). This lack of matching, say between quarters t and t+1, is overcome in the following way: take the parameters estimated using the quarter t+1 hedonic regression and price out all of the housing models (i.e., sales) that appeared in quarter t. This generates *predicted quarter t+1 prices for the quarter t models* , $p_n^{\,t+1}(t)$, as follows:

$$(17)\ p_n^{\,t+1}(t) \equiv \alpha^{t+1*} + \beta^{t+1*} L_n^{\,t} + \gamma^{t+1*}(1 - \delta^{t+1*} A_n^{\,t})S_n^{\,t}\ ; \qquad t = 1,...,13;\ n = 1,...,N(t)$$

where α^{t*}, β^{t*}, γ^{t*} and δ^{t*} are the parameter estimates for the period t regression (16) for t = 1,...,14. Now we have a set of "matched" quarter t+1 prices for the models that

[30] This theory dates back to Court (1939; 108) as his hedonic suggestion number one. His suggestion was followed up by Griliches (1971a; 59-60) (1971b; 6) and Triplett and McDonald (1977; 144). More recent contributions to the literature include Diewert (2003b), Haan (2003) (2009) (2010), Triplett (2004) and Diewert, Heravi and Silver (2009).

[31] Due to the fact that the regressions defined by (15) have a constant term and are essentially linear in the explanatory variables, the sample residuals in each of the regressions will sum to zero. Hence the sum of the predicted prices will equal the sum of the actual prices for each period. Thus the sum of the actual prices in the denominator of (17) will equal the sum of the corresponding predicted prices and similarly, the sum of the actual prices in the numerator of (19) will equal the corresponding sum of the predicted prices.

appeared in period t and we can form the following *Laspeyres type matched model index*, going from quarter t to t+1:

$$(18)\ P_{HIL}(t,t+1) \equiv \sum_{n=1}^{N(t)} 1\ p_n^{t+1}(t) / \sum_{n=1}^{N(t)} 1\ p_n^{t}\ ; \qquad t = 1,...,13.$$

Note that the quantity that is associated with each price is 1; basically, each housing unit is unique and cannot be matched except through the use of a model.

The same method can be used going backwards from the housing sales that took place in quarter t+1; take the parameters for the quarter t hedonic regression and price out all of the housing models that appeared in quarter t+1 and generate predicted prices, $p_n^t(t+1)$ for these t+1 models:

$$(19)\ p_n^t(t+1) \equiv \alpha^{t*} + \beta^{t*} L_n^{t+1} + \gamma^{t*}(1 - \delta^{t*} A_n^{t+1}) S_n^{t+1}\ ; \qquad t = 1,...,13;\ n = 1,...,N(t+1).$$

Now we have a set of "matched" quarter t prices for the models that appeared in period t+1 and we can form the following *Paasche type matched model index*, going from quarter t to t+1:

$$(20)\ P_{HIP}(t,t+1) \equiv \sum_{n=1}^{N(t+1)} 1\ p_n^{t+1} / \sum_{n=1}^{N(t+1)} 1\ p_n^{t}(t+1)\ ; \qquad t = 1,...,13.$$

Once the above Laspeyres and Paasche imputation indexes have been calculated, we can readily form the corresponding *Fisher type matched model index* going from period t to t+1 by taking the geometric average of the two indexes defined by (18) and (20):

$$(21)\ P_{HIF}(t,t+1) \equiv [P_{HIL}(t,t+1) P_{HIP}(t,t+1)]^{1/2}\ ; \qquad t = 1,...,13.$$

The resulting chained Laspeyres, Paasche and Fisher imputation indexes, P_{HIL}, P_{HIP} and P_{HIF}, are plotted below in Figure 5 and are listed in Table 6.

Figure 5: Chained Laspeyres, Paasche and Fisher Imputation Indexes

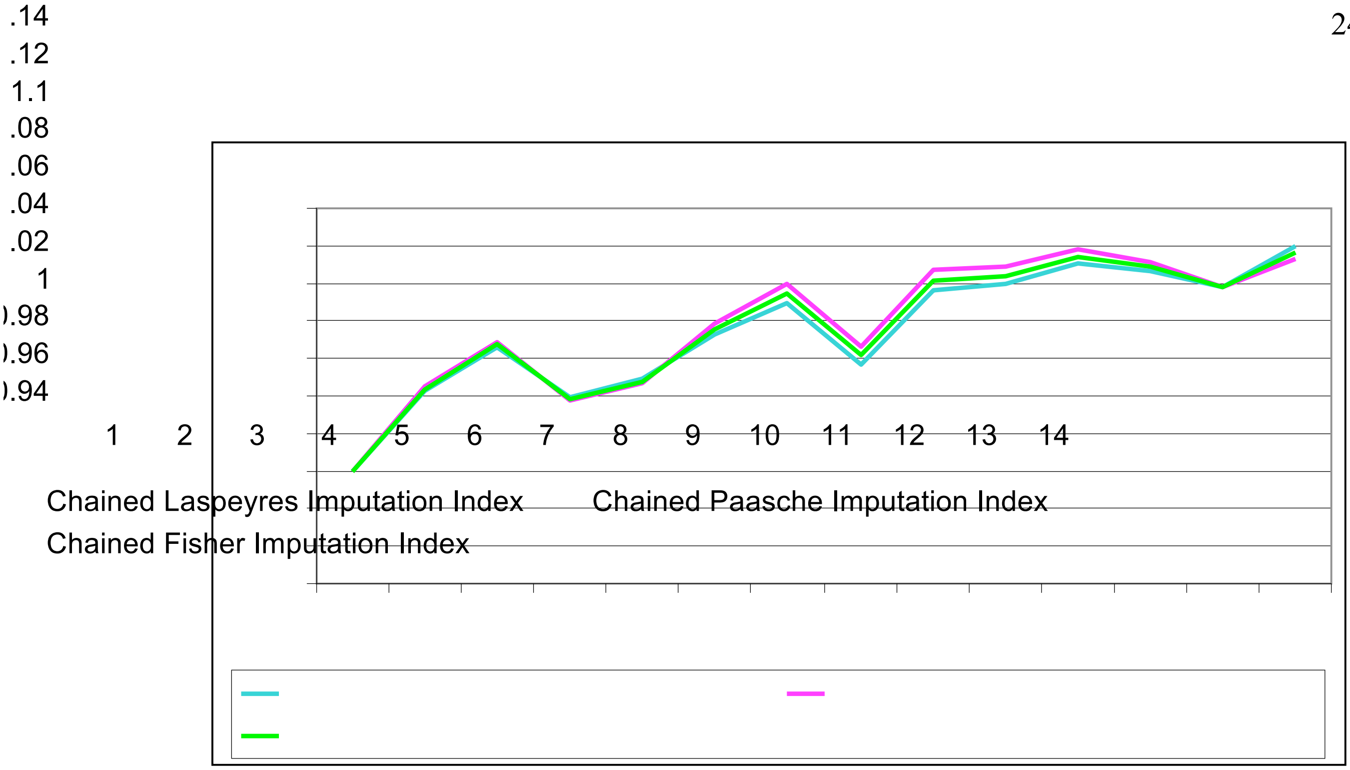


Table 6: Chained Laspeyres, Paasche and Fisher Imputation Indexes

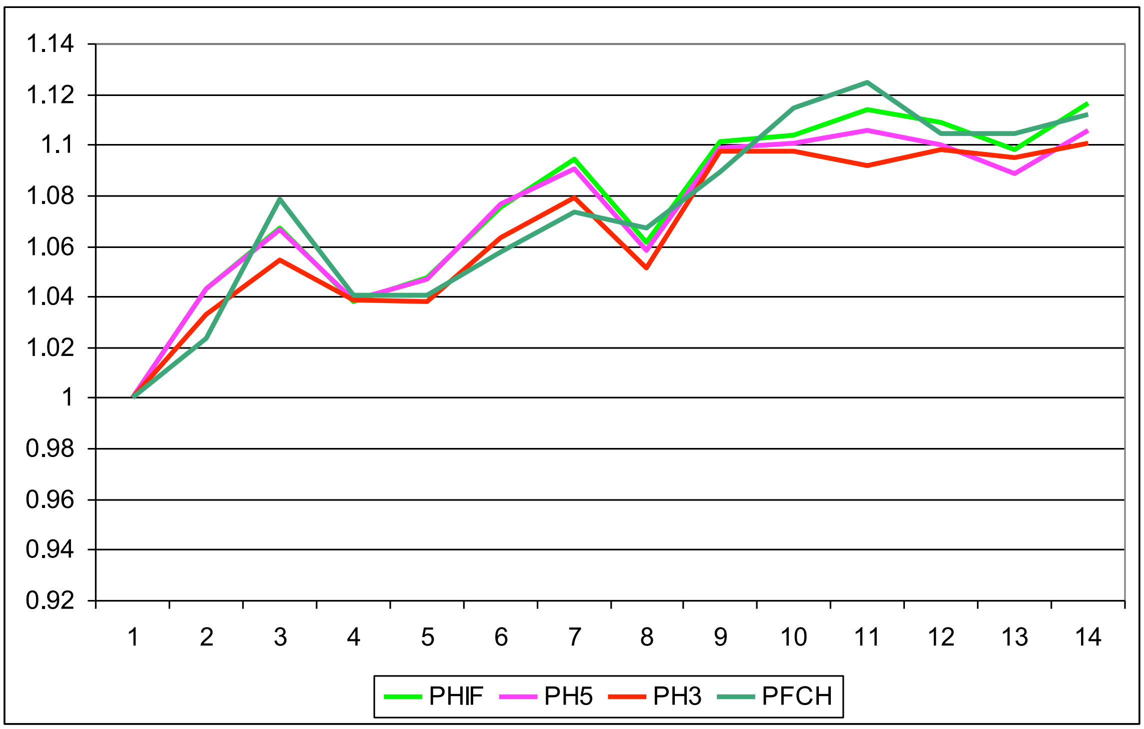
Quarter	P_{HIL}	P_{HIP}	P_{HIF}
1	1.00000	1.00000	1.00000
2	1.04234	1.04479	1.04356
3	1.06639	1.06853	1.06746
4	1.03912	1.03755	1.03834
5	1.04942	1.04647	1.04794
6	1.07267	1.07840	1.07553
7	1.08923	1.10001	1.09460
8	1.05689	1.06628	1.06158
9	1.09635	1.10716	1.10174
10	1.09945	1.10879	1.10411
11	1.11062	1.11801	1.11430
12	1.10665	1.11112	1.10888
13	1.09830	1.09819	1.09824
14	1.11981	1.11280	1.11630

The three imputation indexes are amazingly close.

The Fisher imputation index is our preferred hedonic index thus far; it is better than the time dummy indexes in the previous two sections because the imputation indexes allow the price of land and quality adjusted structures to change independently over time, whereas the time dummy indexes shift the hedonic surface in a parallel fashion. The above empirical results show that the Laspeyres type hedonic imputation index P_{HIL} can provide a very close approximation to the theoretically preferred Fisher type hedonic imputation index P_{HIF}. This is important in the context of producing real time indexes since a reasonably accurate index that covers period t+1 can be constructed using only the period t hedonic regression.

Our two "best" indexes thus far are the Fisher imputation index and the Stratified Chained Fisher index P_{FCH}.[32] These two "best" indexes are plotted in Figure 6 along with the Log Log time dummy indexes P_{H3} and the Linear time dummy index with quality adjusted structures P_{H5}. Note that all of the indexes except P_{H3} indicate downward movements in quarters, 4, 8, 12 and 13 and upward movements in the other quarters (P_{H3} moves up in quarter 12 instead of falling like the other indexes).

Figure 6: The Fisher Hedonic Imputation Price Index P_{HIF}, the Chained Matched Model Stratified Fisher Index P_{FCH}, the Linear Time Dummy Hedonic Regression Index P_{H5} and the Log Log Time Dummy Hedonic Regression Index P_{H3}.



This completes our discussion of basic hedonic regression methods that could be used in order to construct an overall index of house prices. In the following sections, we will study various hedonic regression methods that could be used in order to construct *separate* indexes for the price of housing land and for housing structures.

7. The Construction of Land and Structures Price Indexes: Preliminary Approaches

[32] The stratified sample chained Fisher index that uses imputed prices, P_{IFCH}, is just as good as P_{FCH} on theoretical grounds and in our sample, the two indexes were virtually the same. Overall, the hedonic imputation index P_{HIF} should be preferred to P_{FCH} and P_{IFCH} since the stratified sample indexes will have a certain amount of unit value bias that will probably be greater than any functional form bias in P_{HIF}.

It is reasonable to develop a cost of production approach to the pricing of a newly built house.[33] Thus for a newly built house during quarter t, the total cost of the property after the structure is completed will be approximately equal to the floor space area of the structure, say S square meters, times the building cost per square meter, γ^t say, plus the cost of the land, which will be equal to the cost per square meter, β^t say, times the area of the land site, L. Now think of a sample of newly built properties of the same general type, which have prices p_n^t in quarter t and structure areas S_n^t and land areas L_n^t. The prices of these newly built properties, p_n^t, should be approximately equal to costs of the above type, $\beta^t L_n^t + \gamma^t S_n^t$ plus error terms, which we assume have zero means. This model for pricing the sales of new structures is generalized to include the pricing of used structures by introducing quality adjusted structures in the usual way. This leads to the following hedonic regression model for the entire data set where β^t (the price of land), γ^t (the price of constant quality structures) and δ (the decade depreciation rate) are the parameters to be estimated in the following regression model:[34] [35]

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

Note that a common depreciation rate for all quarters was estimated. Thus the model defined by (22) has 14 unknown β^t parameters, 14 unknown γ^t parameters and one unknown δ or 29 unknown parameters in all. The R^2 for this model was equal to .8847, which is the highest yet for regressions using the entire data set.[36] The log likelihood was −10642.0, which is considerably higher than the log likelihoods obtained for the two time dummy hedonic regressions that used prices as the dependent variable (recall the

---

[33] This additive approach was suggested by several researchers, including 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) and Statistics Portugal (2009). The specific model defined by (22) was suggested by Diewert (2007) and implemented by Diewert, Haan and Hendriks (2010). Thus the model in this section is a *supply side model* as opposed to the *demand side Cobb Douglas model* of McMillen (2003) studied earlier. See Rosen (1974) for a discussion of identification issues in hedonic regression models.

[34] In order to obtain homoskedastic errors, it would be preferable to assume multiplicative errors in equation (22) 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 (22) 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.

[35] Thorsnes (1997; 101) has a related cost of production model. He assumed that instead of equation (22), the value of the property under consideration in period t, p^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 $p^t = \rho^t\ H(L,K) = \rho^t\ [\alpha L^\sigma + \beta K^\sigma]^{1/\sigma}$ 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^*). Our problem with this model is that there is only one independent time parameter ρ^t whereas our model has two, β^t and γ^t for each t, which allow the price of land and structures to vary freely between periods.

[36] The present model is similar in structure to the hedonic imputation model described in the previous section except that this model is more parsimonious; i.e., there is only one depreciation rate in the present model (as opposed to 14 depreciation rates in the imputation model) and there are no constant terms in the present model. The important factor in both models is that the prices of land and quality adjusted structures are allowed to vary independently across time periods.

regressions associated with the construction of P_{H4} and P_{H5}, where the log likelihoods were −10790.4 and −10697.8). The decade straight line estimated depreciation rate was 0.1068 (0.00284).

The model yields an estimated land price for quarter t equal to β^{t*} and the corresponding quantity of land transacted is equal to $L^t \equiv \sum_{n=1}^{N(t)} L_n^{\,t}$. The estimated period t price for a square meter of quality adjusted structures is γ^{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 land price series $\beta^{1*},...,\beta^{14*}$ (rescaled to equal 1 in quarter 1) is the price series P_{L1} which is plotted in Figure 7 and listed in Table 7 below. The constant quality price series for structures $\gamma^{1*},...,\gamma^{14*}$ (rescaled to equal 1 in quarter 1) is the price series P_{S1} which is plotted in Figure 7 and listed in Table 7. Finally, using the price and quantity data on land and constant quality structures for each quarter t, $(\beta^{t*}, L^t, \gamma^{t*}, S^{t*})$ for t = 1,...,14, an overall house price index can be constructed using the Fisher formula. The resulting price series is P_1 which is also plotted in Figure 7 and listed in Table 7 below. For comparison purposes with P_1, the Fisher hedonic imputation index P_{HIF} is also plotted in Figure 7 and listed in Table 7.

Figure 7: 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 P_{HIF}

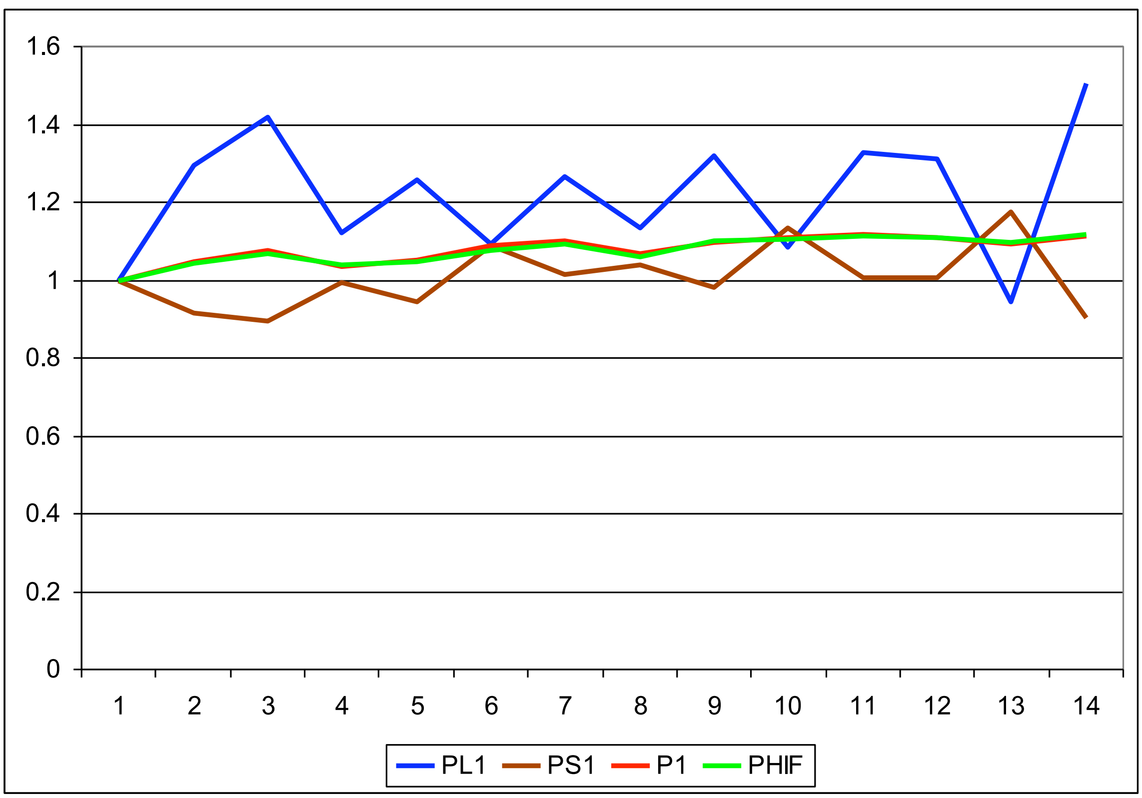


Table 7: The Price of Land P_{L1}, the Price of Quality Adjusted Structures P_{S1}, the Overall Cost of Production House Price Index P_1 and the Fisher Hedonic Imputation House Price Index P_{HIF}

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

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} constructed in the previous section. However, it can also be seen that the price series for land, P_{L1}, and the price series for quality adjusted structures, P_{S1}, are not at all credible: there are large random fluctuation in both series. Note that when the price of land spikes upwards, there is a corresponding dip in the price of structures. This is a sign of multicollinearity between the land and quality adjusted structures variables, which leads to unstable estimates for the prices of land and structures.

There is a tendency for the price of land per meter squared to decrease for large lots. Thus in an attempt to improve upon the results of the hedonic regression model defined by (21), a *linear spline model for the price of land* is implemented.[37] Thus for lots that are less that 160 m^2, we assume that the price of land per meter squared is $\beta_S{}^t$ during quarter t. For sales of properties that have lot sizes between 160 m^2 and 300 m^2, we assume that the cost per m^2 of units of land above 160 m^2 changes to a price of $\beta_M{}^t$ per additional square meter during quarter t. Finally, for large plots of land that are above 300 m^2, we set the marginal price of an additional unit of land above 300 m^2 to equal $\beta_L{}^t$ per square meter during quarter t. For quarter t, let the set of sales n of small, medium and large plots be denoted by $N_S(t)$, $N_M(t)$ and $N_L(t)$ respectively for t = 1,...,14. For sales n of properties that fall into the small land size group during period t, the hedonic regression model is described by (23); for the medium group, by (24) and for the large land size group, by (25):

$$(23)\ p_n{}^t = \beta_S{}^t L_n{}^t + \gamma^t(1 - \delta A_n{}^t)S_n{}^t + \varepsilon_n{}^t\,; \qquad t = 1,...,14;\ n \in N_S(t);$$

[37] This approach follows that of Diewert, Haan and Hendriks (2010). However, the use of linear splines on the size of the lot is due to Francke (2008).

(24) $p_n^t = \beta_S^t[160] + \beta_M^t[L_n^t - 160] + \gamma^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t$; $t = 1,...,14; n\in N_M(t)$;
(25) $p_n^t = \beta_S^t[160] + \beta_M^t[140] + \beta_L^t[L_n^t - 300] + \gamma^t(1 - \delta A_n^t)S_n^t + \varepsilon_n^t$;
$t = 1,...,14; n\in N_L(t)$.

Estimating the model defined by (23)-(25) and using the data for the town of "A", the estimated decade depreciation rate was $\delta^* = 0.1041$ (0.00419). The R^2 for this model was .8875, an increase over the previous no splines model where the R^2 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 3 marginal prices for land were $\beta_S^{1*} = 281.4$ (55.9), $\beta_M^{1*} = 380.4$ (48.5) and $\beta_L^{1*} = 188.9$ (27.5). Thus in quarter 1, the marginal cost per m^2 of small lots is estimated to be 281.4 Euros per m^2. For medium sized lots, the estimated marginal cost is 380.4 Euros/m.^{2} And, for large lots, the estimated marginal cost is 188.9 Euros/m^2. The first period parameter value for quality adjusted structures is $\gamma^{1*} = 978.1$ Euros/m^2 with a standard error of 82.3. The lowest t statistic for all of the 57 parameters is 3.3, so all of the coefficients in this model are significantly different from zero.

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:

(26) $V_{LS}^t \equiv \sum_{n\in N_S(t)} \beta_S^{t*}L_n^t$; $t = 1,...,14$;
(27) $V_{LM}^t \equiv \sum_{n\in N_M(t)} \beta_S^{t*}[160] + \beta_M^{t*}[L_n^t - 160]$; $t = 1,...,14$;
(28) $V_{LL}^t \equiv \sum_{n\in N_L(t)} \beta_S^{t*}[160] + \beta_M^{t*}[140] + \beta_L^{t*}[L_n^t - 300]$; $t = 1,...,14$;
(29) $L_{LS}^t \equiv \sum_{n\in N_S(t)} L_n^t$; $t = 1,...,14$;
(30) $L_{LM}^t \equiv \sum_{n\in N_M(t)} L_n^t$; $t = 1,...,14$;
(31) $L_{LL}^t \equiv \sum_{n\in N_L(t)} L_n^t$ $t = 1,...,14$.

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:

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

The average land prices for small, medium and large lots defined by (32) and the corresponding quantities of land defined by (29)-(31) can be used to form a chained *Fisher land price index*, which we denote by P_{L2}. This index is plotted in Figure 8 and listed in Table 8 below. As in the previous model, the estimated period t price for a square meter of quality adjusted structures is γ^{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 γ^{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 Figure 8 and listed in

Table 8. The *constant quality structures price index* P_{S2} (a normalization of the series $\gamma^{1*},...,\gamma^{14*}$) is also found in Figure 8 and Table 8.

Figure 8: The Price of Land P_{L2}, the Price of Quality Adjusted Structures P_{S2}, the Overall House Price Index P_2 Using Splines on Land and the Chained Stratified Sample Fisher House Price Index P_{FCH}

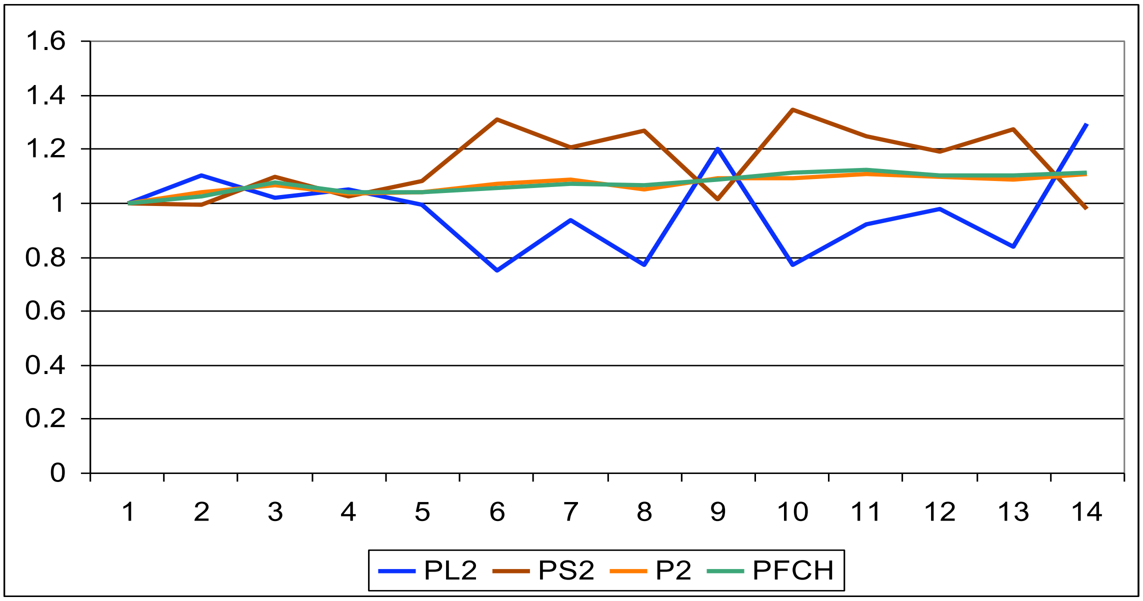


Table 8: The Price of Land P_{L2}, the Price of Quality Adjusted Structures P_{S2}, the Overall House Price Index P_2 Using Splines on Land and the Chained Stratified Sample Fisher House Price Index P_{FCH}

Quarter	P_{L2}	P_{S2}	P_2	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.10534	0.99589	1.04137	1.02396
3	1.02008	1.09803	1.06465	1.07840
4	1.05082	1.02542	1.03608	1.04081
5	0.99379	1.08078	1.04294	1.04083
6	0.74826	1.31122	1.06982	1.05754
7	0.93484	1.20719	1.08912	1.07340
8	0.77202	1.26718	1.05345	1.06706
9	1.19966	1.01724	1.09425	1.08950
10	0.77139	1.34813	1.09472	1.11476
11	0.92119	1.24884	1.10596	1.12471
12	0.97695	1.19188	1.09731	1.10483
13	0.84055	1.27531	1.08811	1.10450
14	1.29261	0.97875	1.10613	1.11189

It can be seen that the overall house price index that results from the spline model, P_2, is very close to the chained Fisher index P_{FCH} that was calculated using the stratification approach. However, the spline model does not generate sensible estimates for the price of land, P_{L2} and the price of structures, P_{S2}: both price indexes are volatile but in opposite directions. As was the case with the previous cost of production model, the present model is subject to a multicollinearity problem.[38]

In the following section, an attempt to cure this volatility problem will be made by imposing monotonicity restrictions on the price movements for land and quality adjusted structures.

8. The Construction of Land and Structures Price Indexes: Approaches Based on Monotonicity Restrictions

It is likely that Dutch construction costs did not fall significantly during the sample period.[39] If this is the case, then these monotonicity restrictions on the quarterly prices of quality adjusted structures, $\gamma^1, \gamma^2, \gamma^3, \ldots, \gamma^{14}$, can be imposed on the hedonic regression model (22)-(24) in the previous section by replacing the constant quality quarter t structures price parameters γ^t by the following sequence of parameters for the 14 quarters: γ^1, $\gamma^1 + (\phi^2)^2$, $\gamma^1 + (\phi^2)^2 + (\phi^3)^2, \ldots$, $\gamma^1 + (\phi^2)^2 + (\phi^3)^2 + \ldots + (\phi^{14})^2$ where $\phi^2, \phi^3, \ldots, \phi^{14}$ are scalar parameters.[40] 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 γ^t in equations (23)-(25) in order to obtain a *linear spline model for the price of land with monotonicity restrictions on the price of constant quality structures*.

Using the data for the town of "A", the estimated decade depreciation rate was $\delta^* = 0.1031$ (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 3 marginal prices for land are $\beta_S^{1*} = 278.6$ (37.2), $\beta_M^{1*} = 380.3$ (41.0) and $\beta_L^{1*} = 188.0$ (21.4) and these estimated parameters are virtually identical to the corresponding parameters in the previous unrestricted model. The first period parameter value for quality adjusted structures is $\gamma^{1*} = 980.5$ (49.9) Euros/m^2 which is little changed from the corresponding unrestricted estimate of 978.1 Euros/m^2.

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

---

[38] Comparing Figures 7 and 8, it can be seen that in Figure 7, the price index for land is *above* the overall price index for the most part while the price index for structures is below the overall index but in Figure 8, this pattern *reverses*. This instability is again an indication of a multicollinearity problem.

[39] Some direct evidence on this assertion will be presented in the following section.

[40] 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 structures prices only.

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

Now use equations (26)-(32) 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 Figure 9 and listed in Table 9 below. As in the previous two models, the estimated period t price for a square meter of quality adjusted structures is γ^{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 γ^{t*} and S^{t*} were combined with the three land price and quantity series to form a chained *overall Fisher house price index* P_3 which is graphed in Figure 9 and listed in Table 9. The *constant quality structures price index* P_{S3} (a normalization of the series $\gamma^{1*},...,\gamma^{14*}$) is also found in Figure 9 and Table 9.

Figure 9: 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 Unrestricted Overall House Price Index Using Splines on Land P_2

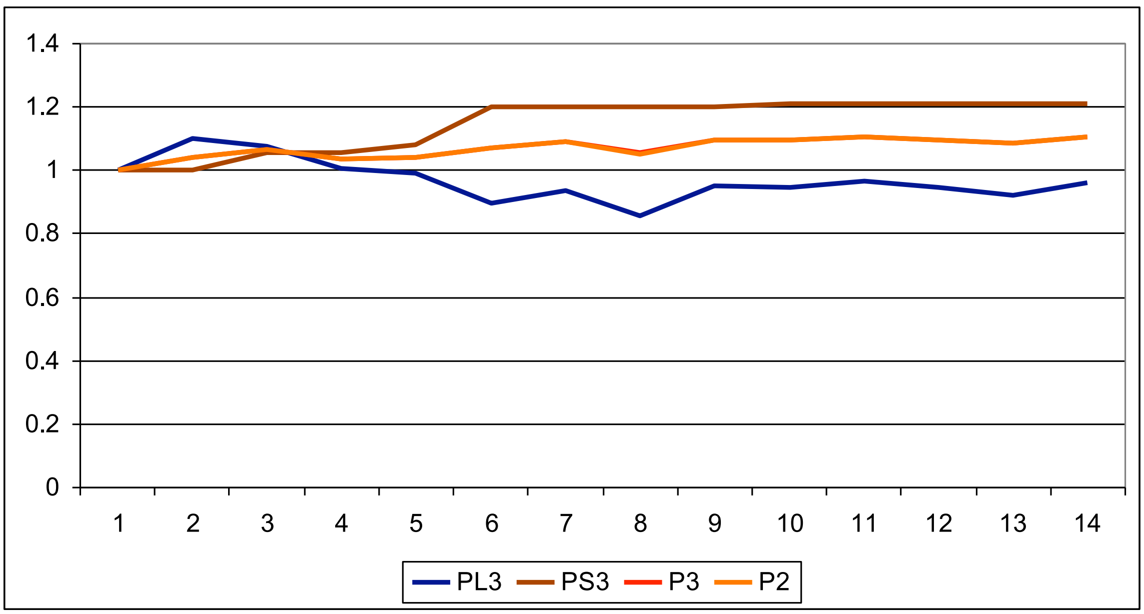


Table 9: The Price of Land P_{L3}, the Price of Quality Adjusted Structures P_{S3}, the Overall House Price Index with Monotonicity Restrictions on Structures P_3 and the Unrestricted Overall House Price Index Using Splines on Land P_2

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

From Figure 9, it can be seen that the new overall house price index P_3 that imposed monotonicity on the quality adjusted price of structures cannot 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. It can also be seen that the new land and structures price indexes look "reasonable"; the fluctuations in the price of land and quality adjusted structures are no longer violent. Finally, we note that the overall index P_3 is quite close to our previously recommended indexes, the matched model stratified chained Fisher index P_{FCH}, and the Fisher hedonic imputation index, P_{HIF}.

Although the above results look "reasonable", the early rapid increase in the price of structures and the slow growth in the index from quarter 6 to 14 looks somewhat unlikely. Thus in the following section, we will try one more method for extracting separate structures and land components out of real estate sales data.

9. The Construction of Land and Structures Price Indexes: An Approach Based on the Use of Exogenous Information on the Price of Structures

Many countries have new construction price indexes available on a quarterly basis. This is the case for the Netherlands.[41] 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 section 6 above.

Recall equations (23)-(25) in section 7 above. These equations are the estimating equations for the unrestricted hedonic regression model based on costs of production. In the present section, the constant quality house price parameters, the γ^t for t =2,...,14 in (23)-(25), are replaced by the following numbers, which involve only the single unknown parameter γ^1:

(34) $\gamma^t = \gamma^1\mu^t$; $\qquad$ t = 2,3,...,14

[41] From the Central Bureau of Statistics (2010) online source, Statline, the following series was downloaded for the New Dwelling Output Price Index for the 14 quarters in our sample of house sales in "A": 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}.

where μ^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. The new hedonic regression model is again defined by equations (23)-(25) except that the 14 unknown γ^t parameters are now assumed to be defined by (34), so that only γ^1 needs to be estimated for this new model. Thus the number of parameters to be estimated in this new restricted model is 44 as compared to the old number, which was 57.

Using the data for the town of "A", the estimated decade depreciation rate was $\delta^* = 0.1028$ (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 7, which was .8875. The log likelihood was −10640.1, a decrease of 10 over the previous monotonicity restricted model. The first period parameter values for the 3 marginal prices for land are $\beta_S{}^{1*} = 215.4$ (30.0), $\beta_M{}^{1*} = 362.6$ (46.7) and $\beta_L{}^{1*} = 176.4$ (28.4). These new estimates differ somewhat from our previous estimates for these parameters. The first period parameter value for quality adjusted structures is $\gamma^{1*} = 1085.9$ (22.9) Euros/m^2 which is substantially changed from the corresponding unrestricted estimate which is 980.5 Euros/m^2. 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 (26)-(32) in order to construct a chained Fisher index of land prices, which we denote by P_{L4}. This index is plotted in Figure 10 and listed in Table 10 below. As for the previous three models, the estimated period t price for a square meter of quality adjusted structures is γ^{t*} (which in turn is now equal to $\gamma^{1*}\mu^t$) 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 γ^{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 Figure 10 and listed in Table 10. The *constant quality structures price index* P_{S4} (a normalization of the series $\gamma^{1*},...,\gamma^{14*}$) is also found in Figure 10 and Table 10.

Figure 10: 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

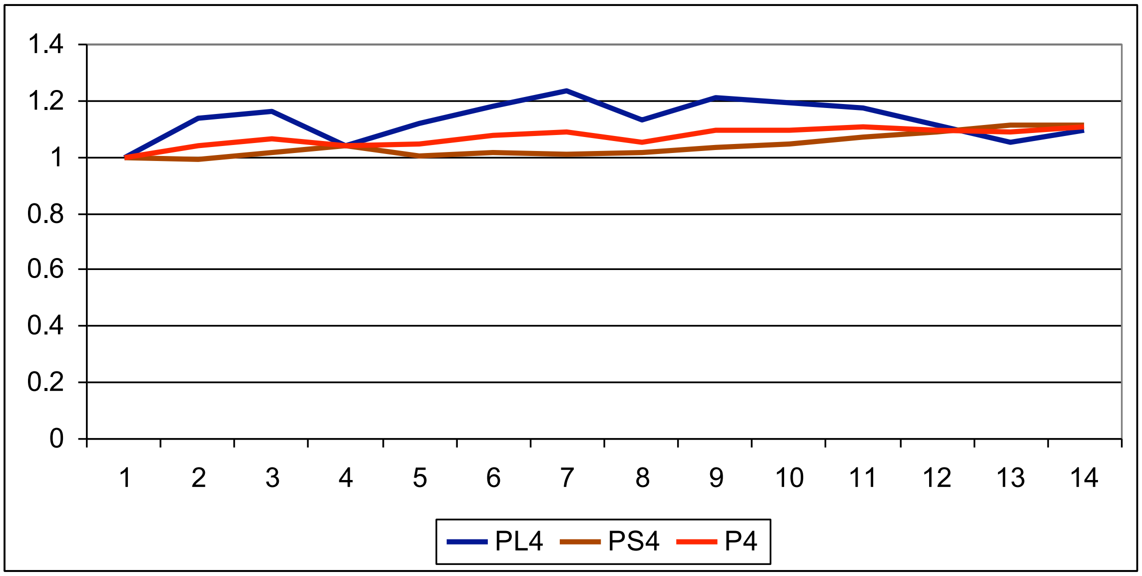


Table 10: The Price of Land P_{L4}, the Price of Quality Adjusted Structures P_{S4}, and the Overall House Price Index using Exogenous Information on the Price of Structures P_4

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

Comparing Figures 9 and 10, it can be seen that the imposition of the national growth rates for new dwelling construction costs has totally changed the nature of our land and structures price indexes: in Figure 9, the price series for land lies below the overall house price series for most of the sample period while in Figure 10, the pattern is reversed as 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). Again, this is a reflection of the large amount of variability in the data and the multicollinearity between selling price, the quantity of land and the quantity of structures.

Which model is best? It is difficult to be definitive at this stage: on statistical grounds, the log likelihood is somewhat higher for the previous model that generated the P_3 overall index (and thus it should be preferred from this point of view) but the pattern of price changes for land and structures seems more believable for the present model using exogenous information on structures prices (and thus the exogenous information model should be preferred).

We conclude this section by listing and charting our four preferred overall indexes. These four indexes are the matched model chained Fisher stratified sample index P_{FCH}[42] studied in section 2, the chained Fisher hedonic imputation index P_{HIF} studied in section 6, the index P_3 that resulted from the cost based hedonic regression model with monotonicity restrictions studied in section 8 and the index P_4 that was generated by the cost based hedonic regression model which used exogenous information on the price of structures studied in the present section. As can be seen from Figure 11 below, all four of these indexes paint much the same picture. Note that P_3 and P_4 are virtually identical.

Table 11: House Price Indexes Using Exogenous Information P_4 and Using Monotonicity Restrictions P_3, the Fisher Chained Imputation Index P_{HIF} and the Chained Fisher Stratified Sample Index P_{FCH}

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

Figure 11: House Price Indexes Using Exogenous Information P_4 and Using Monotonicity Restrictions P_3, the Fisher Chained Imputation Index P_{HIF} and the Chained Fisher Stratified Sample Index P_{FCH}

[42] P_{IFCH} is equally preferred to P_{FCH} but since it is so close to P_{FCH}, it is not listed.

All things considered, the hedonic imputation index P_{HIF} is our preferred index (since it has fewer restrictions than the other indexes and seems closest to a matched model index in spirit) followed by the two cost of production hedonic indexes P_4 and P_3 followed by the stratified sample indexes P_{FCH} and P_{IFCH} (which are likely to have some unit value bias).[43] If separate land and structures indexes are required, then the cost based hedonic regression model that used exogenous information on the price of structures is our preferred model.

A problem with the hedonic regression models discussed in sections 4, 5 and in 7-9 is that as the data for a new quarter are added, the old index values presumably will change as well when a new hedonic regression is run with the additional data. This problem is addressed in the next section.

10. Rolling Window Hedonic Regressions

Recall the last hedonic regression model that was discussed in the previous section. This model was defined by equations (23)-(25) and (34), where equations (34) imposed exogenous information on the price of structures over the sample period. A problem with this hedonic regression model (and all the other hedonic regression models discussed in

---

[43] However, the hedonic regression based indexes can be biased as well if important explanatory variables are omitted and if an “incorrect” functional form for the hedonic regression is chosen. But in general, hedonic regression methods are probably preferred over stratification methods.

this paper with the exception of the hedonic imputation models) is that when more data are added, the indexes 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.[44] 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, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010) in their hedonic regression models for Tokyo house prices.[45]

We implement the rolling window procedure for the last model in the previous section with a window length of 9 quarters. Thus the initial hedonic regression model defined by (23)-(25) and (34) is implemented for the first 9 quarters. The resulting indexes for the price of land, constant quality structures and the overall index are denoted by P_{RWL4}, P_{RWS4} and P_{RW4} respectively and are listed in the first 9 rows of Table 12 below.[46] Next a regression covering the data for quarters 2-10 was run and the land, structures and overall price indexes generated by this model were used to update the initial indexes in the first 9 rows of Table 12; i.e., the price of land in quarter 10 of Table 12 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) that was obtained from the second 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. The rolling window indexes can be compared to their one big regression counterparts (the model in the previous section) by looking at Table 12. Recall that the estimated depreciation rate and the estimated Quarter 1 price of quality adjusted structures for the last model in the previous section were $\delta^* = 0.1028$ and $\gamma^{1*} = 1085.9$ respectively. If by chance, the six rolling window hedonic regressions each generate the same estimates for δ and γ, then the indexes generated by the rolling window regressions

[44] Users may tolerate a few revisions to recent data but typically, users would not like all the numbers to be revised back into the indefinite past as new data become available.

[45] An analogous procedure has also been recently 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 to the scanner data context.

[46] 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 10, P_{S4}.

would coincide with the indexes P_{L4}, P_{S4} and P_4 that were described in the previous section. The six estimates for δ generated by the six rolling window regressions are 0.10124, 0.10805, 0.11601, 0.11103, 0.10857 and 0.10592. The six estimates for γ^1 generated by the six 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 P_4 estimates of 0.1028 and 1085.9, they are fairly close and so we can expect the rolling window indexes to be fairly close to their counterparts for the last model in the previous section. The R^2 values for the six rolling window regressions were .8803, .8813, .8825, .8852, .8811 and .8892.

Table 12: 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

The rolling window series for the price of quality adjusted structures, P_{RWS}, is not listed in Table 12 because it is identical to the series P_{S4}, which was described in the previous section.[47] It can be seen that the new rolling window price series for land, P_{RWL}, is extremely close to its counterpart in the previous section, P_{L4}, and the overall rolling window price series for detached dwellings in "A", P_{RW}, is also close to its counterpart in the previous section, P_4. These series are so close to each other that a chart shows practically no differences, which explains why we have not provided a chart for the series in Table 12.

Our conclusion here is that rolling window hedonic regressions can give pretty much the same results as a longer hedonic regression that covers the sample period. Thus the use of rolling window hedonic regressions can be recommended for statistical agency use.

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

A final topic of interest is: how can the results of hedonic regression models for sales of houses be adapted to give estimates for a price index for the stock of houses? This topic is briefly addressed in the following section.

11. The Construction of Price Indexes for the Stock of Dwelling Units using the Results of Hedonic Regressions on the Sales of Houses

In this section, we will show how the hedonic regression models estimated in sections 6 and 9 can be used in order to form approximate price indexes for the stock of dwelling units.

Recall that the system of hedonic regression equations for the hedonic imputation model discussed in section 6 was equations (15), where L_n^t, S_n^t and A_n^t denote, respectively, the land area, structure area, and age (in decades) of the detached house n that was sold in period t. In order to form a price index for the stock of dwelling units in the town of "A", it would be necessary to know L, S and A for the entire stock of detached houses in "A" for some base period. This information is not available to us but we treat the total number of houses sold over the 14 quarters as an approximation to the stock of dwellings of this type.[48] Thus there are $N \equiv N(1) + N(2) + ... + N(14) = 2289$ houses that were transacted during the 14 periods in our sample.[49]

Recall the hedonic regression equations (15) in section 6 and let α^{t*}, β^{t*}, γ^{t*} and δ^{t*} denote the estimates for the unknown parameters in (15) for quarter t for t = 1,...,14. Our approximation to *the total value of the housing stock for quarter t*, V^t, is defined as follows:

$$(35)\ V^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} [\alpha^{t*} + \beta^{t*}L_n^s + \gamma^{t*}(1 - \delta^{t*}A_n^s)S_n^s]\ ; \qquad t = 1,...,14.$$

Thus V^t is simply the imputed value of all of the houses that traded during the 14 quarters in our sample using the estimated regression coefficients for the quarter t hedonic imputation regression 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". This is a form of a Lowe index; see the CPI Manual (ILO et al. 2004) for additional material on the properties of Lowe indexes. This price index for the stock of housing units is compared with the corresponding Fisher hedonic imputation price index, P_{HIF} from section 6, in Table 13 and Figure 12 below.

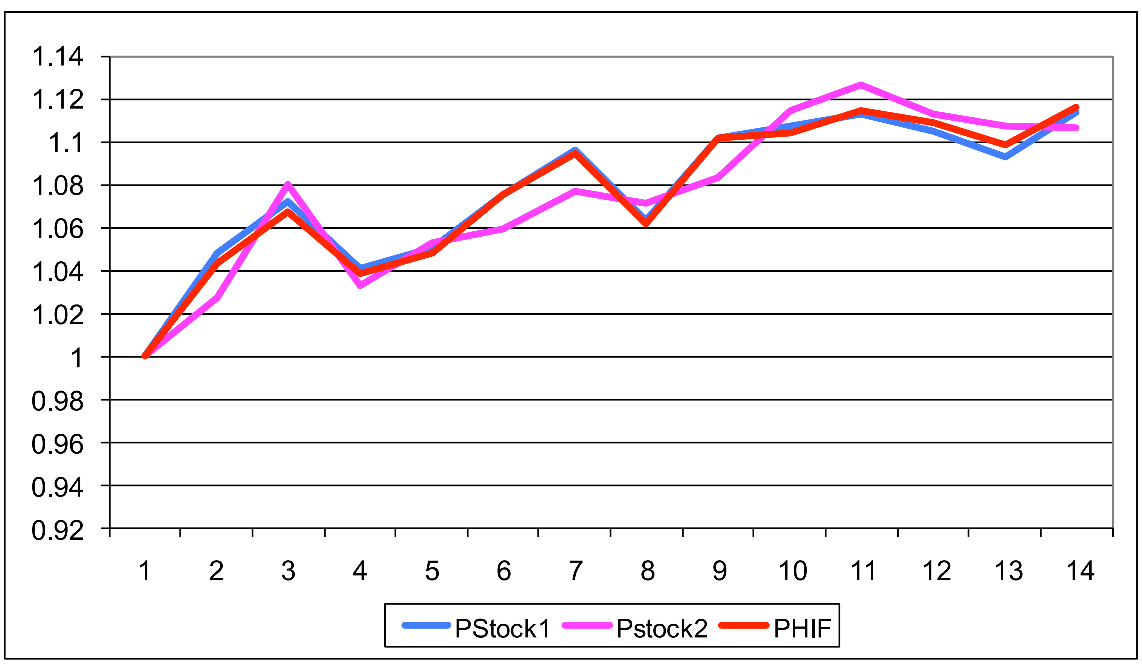
[48] This approximation would probably be an adequate one if the sample period were a decade or so. Obviously, our sample period of 14 quarters is too short to be a good approximation but the method we are suggesting can be illustrated using this rough approximation. There are also sample selectivity problems with this approximation; i.e., new houses will be over represented using this method.

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

Table 13: Approximate Stock Price Indexes P_{Stock1} and P_{Stock2} Based on Hedonic Imputation and on Stratification and the Fisher Hedonic Imputation Sales Price Index P_{HIF}

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

Figure 12: Approximate Stock Price Indexes P_{Stock1} and P_{Stock2} Based on Hedonic Imputation and on Stratification and the Fisher Hedonic Imputation Sales Price Index P_{HIF}



It can be seen that the differences between the approximate *stock* house price series based on the hedonic imputation regressions, P_{Stock1}, and the corresponding hedonic imputation index for *sales* of houses based on the hedonic regression explained in section 6, P_{HIF}, are generally quite small, less than one half of a percentage point for each quarter.

For comparison purposes, an additional *approximate stock price index based on the stratification model* explained in section 2, P_{Stock2}, is listed in Table 13 and graphed in Figure 12. This index uses the positive unit value cell prices for the nonempty cells in each quarter in our stratification scheme and it uses the imputed prices based on the hedonic imputation regressions for each quarter for the empty cells in each quarter. The quantity vector used for P_{Stock2} is the sample total quantity vector by cell and thus P_{Stock2} is an alternative Lowe 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 indexes for some observations. These differences are probably due to the existence of some unit value bias in the stratification indexes. Thus while stratification indexes can be constructed for the stock of dwelling units of a certain type and location (with the help of hedonic imputation regressions), it appears that the resulting stock indexes will not be as accurate as indexes that are entirely based on the use of hedonic regressions.[50]

The same kind of construction of a stock index can be done for the other hedonic regression models that were implemented for sales of houses in previous sections. We will conclude this section by constructing an approximate stock price index using the results of the cost based hedonic regression model that used exogenous information on the price of structures that was explained in section 9 above. Recall that this model was defined by equations (23)-(25) and (34). Recall also that the sets of period t sales of small, medium and large lot houses were defined as $N_S(t)$, $N_M(t)$ and $N_L(t)$ respectively and the total number of sales in period t was defined as N(t) for t = 1,...,14. Denote the estimated parameter values for the model (23)-(25) and (34) by δ^*, γ^{1*}, and β_S^{t*}, β_M^{t*}, β_L^{t*} 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, V_{LL}^t for t = 1,...,14, are defined by (36)-(38) respectively:

$$(36)\ V_{LS}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_S(r)} \beta_S^{t*} L_n^r\ ; \qquad t = 1,...,14;$$

$$(37)\ V_{LM}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_M(r)} \{\beta_S^{t*}[160] + \beta_M^{t*}[L_n^r - 160]\}\ ; \qquad t = 1,...,14;$$

$$(38)\ V_{LL}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_L(r)} \{\beta_S^{t*}[160] + \beta_M^{t*}[140] + \beta_L^{t*}[L_n^r - 300]\}\ ; \qquad t = 1,...,14;$$

$$(39)\ V_S^t \equiv \sum_{r=1}^{14} \sum_{n=1}^{N(r)} \gamma^{1*} \mu^t (1 - \delta^* A_n^r) S_n^r\ ; \qquad t = 1,...,14.$$

[50] If the imputed prices described in section 2 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_{Stock2}) 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 estimated period t value of quality adjusted structures, V_S^t, is defined by (39) above, where all structures traded during the 14 quarters are included in this imputed total value. The *quantities* that correspond to the above period t valuations of the stock of structures and the 3 land stocks are defined as follows:[51]

$$(40)\ Q_{LS}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_S(r)} L_n^r\ ; \qquad t = 1,...,14;$$

$$(41)\ Q_{LM}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_M(r)} L_n^r\ ; \qquad t = 1,...,14;$$

$$(42)\ Q_{LL}^t \equiv \sum_{r=1}^{14} \sum_{n \in N_L(r)} L_n^r \qquad t = 1,...,14;$$

$$(43)\ Q_S^t \equiv \sum_{r=1}^{14} \sum_{n=1}^{N(r)} (1 - \delta^* A_n^r) S_n^r\ ; \qquad t = 1,...,14.$$

The *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 (36)-(43) are defined in the usual way:

$$(44)\ P_{LS}^t \equiv V_{LS}^t/Q_{LS}^t\ ;\ P_{LM}^t \equiv V_{LM}^t/Q_{LM}^t\ ;\ P_{LL}^t \equiv V_{LL}^t/Q_{LL}^t\ ;\ P_S^t \equiv V_S^t/Q_S^t\ ; \qquad t = 1,...,14.$$

With prices defined by (44) and quantities defined by (40)-(43), an approximate stock index of land prices, P_{LStock}, is formed by aggregating the three types of land and an overall approximate stock index of house prices, P_{Stock}, is formed by aggregating the three types of land with the constant quality structures. Since quantities are constant over all 14 quarters, the Laspeyres, Paasche and Fisher indexes are all equal.[52] An approximate constant quality stock price for structures, P_{SStock}, is formed by normalizing the series P_S^t. The approximate stock price series, P_{LStock}, P_{SStock} and P_{Stock} are listed in Table 14 and are charted in Figure 13 below. For comparison purposes, the corresponding price indexes based on sales of properties for the model presented in section 9, P_{L4}, P_{S4} and P_4, are also listed in Table 14.

Table 14: Approximate Price Indexes for the Stock of Houses P_{Stock}, the Stock of Land P_{LStock}, the Stock of Structures P_{SStock} and the Corresponding Sales Indexes P_{L4}, P_{S4} and P_4.

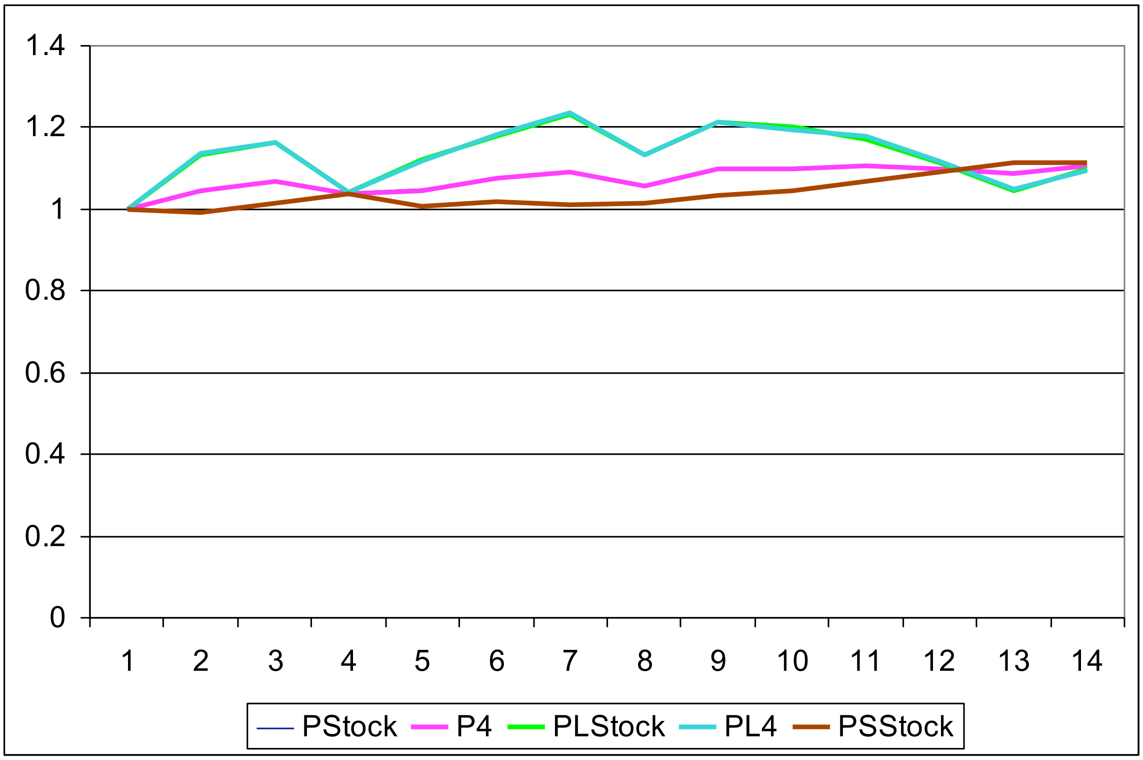
Quarter	P_{Stock}	P_4	P_{LStock}	P_{L4}	P_{SStock}	P_{S4}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.04331	1.04373	1.13279	1.13864	0.99291	0.99291
3	1.06798	1.06752	1.16171	1.16526	1.01518	1.01518
4	1.04042	1.03889	1.04209	1.04214	1.03947	1.03947
5	1.04767	1.04628	1.11973	1.11893	1.00709	1.00709
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

[51] The quantities defined by (39)-(42) are constant over the 14 quarters: $Q_{LS}^t = 77455$, $Q_{LM}^t = 258550$, $Q_{LL}^t = 253590$ and $Q_S^t = 238476.3$ for t = 1,...,14.

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

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

Figure 13: Approximate Price Indexes for the Stock of Houses P_{Stock}, the Stock of Land P_{LStock}, the Stock of Structures P_{SStock} and the Corresponding Sales Indexes P_{L4} and P_4.



From Table 14, it can be seen that the new stock price index for structures, P_{SStock}, coincides with the sales type price index for constant quality structures, P_{S4}, that was described in section 9 above. Thus P_{S4} is not charted in Figure 13. From Figure 13, it can be seen that the overall approximate price index for the stock of houses in "A", P_{Stock}, cannot be distinguished from the corresponding overall sales price index P_4 that was discussed in section 9 and similarly, the overall approximate price index for the stock of land in "A", P_{LStock}, cannot be distinguished from the corresponding overall sales price index for land in "A", P_{L4}. However, Table 14 shows that there are small differences between the stock and sales indexes.

Our conclusion here is that the hedonic regression models for the sales of houses can readily be adapted to yield Lowe type price indexes for the stocks of houses and generally, there do not appear to be major differences between the two index types.

12. Conclusion

Several tentative conclusions can be drawn from this study:

- If information on the sales of houses during a quarter or month is available by location and if information on the age of the houses, the type of housing and their living space and lot size areas is also available, then stratification methods and hedonic regression methods for constructing house price indexes of sales will give approximately the same answers, provided the information on age, lot size and house size is used for both types of method.
- Our preferred method for constructing a sales price index is the hedonic imputation method explained in section 6 but virtually all forms of hedonic regression model using the three main characteristics used in this study give much the same answer, at least when the target index is an overall house price index.
- However, when a linear specification based on a cost of production approach to hedonic regressions is used, the fit to the data is usually considerably better than the fits that result when alternative hedonic regression models are used.
- Rolling year indexes can be used to eliminate seasonality or traditional econometric methods can be applied to the unadjusted house price series; see section 3 above.
- A problem with many hedonic regression models for house prices is that as new data become available, the historical series must constantly be revised. However, if the rolling window technique pioneered by Shimizu, Nishimura and Watanabe (2011) is used, this problem is solved and the results do not differ materially from the one big regression approach that leads to constant revisions; see section 10.
- If separate land and structures house price indexes are required, then the methods based on the cost of production approach with restrictions seem promising; see the method based on imposing monotonicity restrictions on the price of structures explained in section 8 and the method based on the use of exogenous information on the price of structures explained in section 9.
- Hedonic regression methods based on the sales of dwelling units can readily be adapted to yield price indexes for the stock of dwellings; see section 11.

Of course, this is only one study and the results here need to be confirmed using other data sets. However, it seems likely that at least some of the above conclusions will not be overturned by future research.

Some problems that require future research are:

- The techniques here need to be extended to encompass the use of additional characteristics.
- It would be useful to extend the spline treatment of plot size to the size of the structure; i.e., it is likely that the price per meter squared of structure increases as the structure size increases and a spline model could capture this variation.

- The basic method used here that concentrated on holding location constant and using information on three main detached house characteristics needs to be adapted to deal with sales of apartments and row houses, where other characteristics are likely to be important.

References

Alterman, W., W.E. Diewert and R.C. Feenstra (1999), *International Trade Price Indexes and Seasonal Commodities*, Department of Labor, Bureau of Labor Statistics, Washington, D.C.: U.S. Government Printing Office.

Balk, B.M. (1998), "On the Use of Unit Value Indices as Consumer Price Subindices", in: Lane, W. (Ed.) *Proceedings of the Fourth Meeting of the International Working Group on Price Indices*, Washington, DC: Bureau of Labour Statistics.

Balk, B.M. (2008), *Price and Quantity Index Numbers*, New York: Cambridge University Press.

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

Central Bureau of Statistics (2010), "New Dwelling Output Price Indices, Building Costs, 2005 = 100, Price Index: Building Costs including VAT", October 15, Den Haag: Statline, CBS.

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

Court, A. T. (1939), "Hedonic Price Indexes with Automotive Examples", pp. 98-117 in *The Dynamics of Automobile Demand*, New York: General Motors Corporation.

Davis, M.A. and J. Heathcote (2007), "The Price and Quantity of Residential Land in the United States", *Journal of Monetary Economics* 54. 2595-2620.

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, D. Usher (ed.), Chicago: The University of Chicago Press.

Diewert, W.E. (1983), "The Treatment of Seasonality in a Cost of Living Index", pp. 1019-1045 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Diewert, W.E. (1998), "High Inflation, Seasonal Commodities and Annual Index Numbers", *Macroeconomic Dynamics* 2, 456-471.

Diewert, W.E. (1999), "Index Number Approaches to Seasonal Adjustment", *Macroeconomic Dynamics* 3, 1-21.

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

Diewert, W.E. (2003b), "Hedonic Regressions: A Review of Some Unresolved Issues", paper presented at the 7th Meeting of the Ottawa Group, Paris, May 27-29.

http://www.ottawagroup.org/pdf/07/Hedonics%20unresolved%20issues%20-%20Diewert%20(2003).pdf

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., Y. Finkel and Y. Artsev (2009), "Empirical Evidence on the Treatment of Seasonal Products: The Israeli Experience", pp. 53-78 in *Price and Productivity Measurement: Volume 2; Seasonality*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Diewert, W.E., J. de Haan and R. Hendriks (2010), "The Decomposition of a House Price Index into Land and Structures Components: A Hedonic Regression Approach", Discussion Paper 10-01, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.

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. and P. von der Lippe (2010), "Notes on Unit Value Bias", *Journal of Economics and Statistics*, forthcoming.

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

Francke, M.K. (2008), "The Hierarchical Trend Model", pp. 164-180 in *Mass Appraisal Methods: An International Perspective for Property Valuers*, T. Kauko and M. Damato (eds.), Oxford: Wiley-Blackwell.

Francke, M.K. and G.A. Vos (2004), "The Hierarchical Trend Model for Property Valuation and Local Price Indices", *Journal of Real Estate Finance and Economics* 28:2/3, 179-208.

Griliches, Z. (1971a), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change", pp. 55-87 in *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press.

Griliches, Z. (1971b), "Introduction: Hedonic Price Indexes Revisited", pp. 3-15 in *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press.

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 (2003), "Time Dummy Approaches to Hedonic Price Measurement", Paper presented at the Seventh Meeting of the International Working Group on Price Indices, (Ottawa Group), May 27-29, 2003, INSEE, Paris. http://www.insee.fr/en/nom_def_met/colloques/ottawa/ottawa_papers.htm

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*, forthcoming.

Hill, T.P. (1988), "Recent Developments in Index Number Theory and Practice", *OECD Economic Studies* 10, 123-148.

Hill, T.P. (1993), "Price and Volume Measures", pp. 379-406 in *System of National Accounts 1993*, Eurostat, IMF, OECD, UN and World Bank, Luxembourg, Washington, D.C., Paris, New York, and Washington, D.C.

ILO, IMF, OECD, UNECE, Eurostat and World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, ed. by P. Hill, ILO: Geneva.

IMF, ILO, OECD, Eurostat, UNECE and the World Bank (2004), *Producer Price Index Manual: Theory and Practice*, Paul Armknecht (ed.), Washington: International Monetary Fund.

IMF, ILO, OECD, UNECE and World Bank (2009), *Export and Import Price Index Manual*, ed. by M. Silver, IMF: Washington, D.C.

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

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.

Laspeyres, E. (1871), "Die Berechnung einer mittleren Waarenpreissteigerung", *Jahrbücher für Nationalökonomie und Statistik* 16, 296-314.

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.

Mudgett, B.D. (1955), "The Measurement of Seasonal Movements in Price and Quantity Indexes", *Journal of the American Statistical Association* 50, 93-98.

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

Paasche, H. (1874), "Über die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen", *Jahrbücher für Nationalökonomie und Statistik* 12, 168-178.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy* 82, 34-55.

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.

Silver, M. (2009a), "Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices?", IMF Staff Papers 56, 297-322. Washington D.C.: IMF.

Silver, M. (2009b), "Unit Value Indices" Chapter 2 in IMF, ILO, OECD, UNECE and World Bank (2008), *Export and Import Price Index Manual*, ed. by M. Silver, IMF: Washington, D.C.

Silver, M. (2010), "The Wrongs and Rights of Unit Value Indices", *Review of Income and Wealth*, Series 56, Special Issue 1, S206-S223.

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.

Stone, R. (1956), *Quantity and Price Indexes in National Accounts*, Paris: OECD.

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.

Triplett, J.E. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, STI Working Paper 2004/9, OECD Directorate for Science, Technology and Industry, DSTI/DOC(2004)9, Paris: OECD.

Triplett, J.E. and R.J. McDonald (1977), "Assessing the Quality Error in Output Measures: The Case of Refrigerators", *The Review of Income and Wealth* 23:2, 137-156.