Hedonic Imputation versus Time Dummy Hedonic Indexes

Erwin Diewert, Saeed Heravi and Mick Silver
(with a commentary by Jan de Haan)

Discussion Paper 007-07,
Department of Economics,
University of British Columbia,
#997-1873 East Mall,
Vancouver, BC V6T 1Z1 Canada.
Email: diewert@econ.ubc.ca

Abstract:
Statistical offices try to match item models when measuring inflation between two periods. However, for product areas with a high turnover of differentiated models, the use of hedonic indexes is more appropriate since they include the prices and quantities of unmatched new and old models. The two main approaches to hedonic indexes are hedonic imputation (HI) indexes and dummy time hedonic (HD) indexes. This study provides a formal analysis of the difference between the two approaches for alternative implementations of an index that uses weighting that is comparable to the weighting used by the Törnqvist superlative index in standard index number theory. This study shows exactly why the results may differ and discusses the issue of choice between these approaches. An illustrative study for desktop PCs is provided.

Keywords: Hedonic regressions; hedonic indexes; consumer price indexes; superlative indexes.

JEL classification: C43, C82, E31.

1. Introduction

The purpose of this paper is to compare two main and quite distinct approaches to the measurement of hedonic price indexes: time dummy hedonic indexes and hedonic imputation indexes. Both approaches not only correct price changes for changes in the quality of items purchased, but also allow the indexes to incorporate matched and unmatched models. They provide a means by which price change can be measured in product markets where there is a rapid turnover of differentiated models. However, they can yield quite different results. This paper provides a formal exposition of the factors underlying such differences and the implications for choice of method. We consider both weighted and unweighted hedonic regression models. Unweighted hedonic regression models will be considered in sections 2 and 3 below. These models are of course useful in a sampling context where information on the quantity or value of sales (or purchases) is unavailable. Weighted hedonic regression models are considered in sections 4 and 5 below. The weighting is chosen so that if we are actually in a matched model situation for the two periods being considered, then the resulting hedonic regression measures of price change resemble standard superlative index number formulae.

The standard way price changes are measured by national statistical offices is through the use of a matched models methodology. Using this methodology, the details and prices of a representative selection of items are collected in a base reference period and their matched prices collected in successive periods so that the prices of ‘like’ are compared with ‘like’. However, if there is a rapid turnover of available models, then the sample of product prices used to measure price changes

---

1 Our thanks to Ernst Berndt, John Greenlees, Jan de Haan, Alice Nakamura and Jack Triplett for helpful comments. The authors are at the University of British Columbia, the University of Cardiff and the International Monetary Fund respectively. The paper is forthcoming in Price Index Concepts and Measurement, W.E. Diewert, J. Greenlees and C. Hulten, University of Chicago Press.
becomes unrepresentative of the category as a whole. This is as a result of both new unmatched models being introduced (but not included in the sample), and older unmatched models being retired (and thus dropping out of the sample). Hedonic indexes use matched and unmatched models and in doing so put an end to the matched models sample selection bias.\textsuperscript{2} The need for hedonic indexes can be seen in the context of the need to reduce bias in the measurement of the U.S. consumer price index (CPI), which has been the subject of three major reports: the Stigler (1961) Committee Report, the Boskin (1996) Commission Report and the Schultzze and Mackie (2002) Committee on National Statistics Panel Report. Each found the inability to properly remove the effect on price changes of changes in quality to be a major source of bias. Hedonic regressions were considered to be the most promising approach to control for such quality changes, though the Schultzze panel cautioned for the need for further research on methodology:

“Hedonic techniques currently offer the most promising approach for explicitly adjusting observed prices to account for changing product quality. But our analysis suggests that there are still substantial unresolved econometric, data, and other measurement issues that need further attention.” C.L. Schultzze and C. Mackie (2002; 6).

At first sight the two approaches to hedonic indexes appear quite similar. Both rely on hedonic regression equations to remove the effects on price of quality changes. They can also incorporate a range of weighting systems and can be formulated as a geometric, harmonic or arithmetic aggregator function of quality adjusted prices, and as chained or direct, fixed base comparisons. Yet they can give quite different results, even when using comparable weights, functional forms and the same method of making comparisons over periods. This is due to the fact that they work on different averaging principles. The dummy variable method constrains hedonic regression parameters to be the same over time. A hedonic imputation index conversely allows the quality adjustment parameters to change in each period and undertakes two sets of quality adjustments to prices for each comparison of prices between two periods and then averages over these two comparisons.

There has been some valuable research on the two approaches\textsuperscript{3} though no formal analysis has been presented, to the authors’ knowledge with a few exceptions\textsuperscript{4}, of the factors governing the differences between the approaches. Berndt and Rappaport (2001) and Pakes (2003) have highlighted the fact that the two approaches can give different results and both of these papers advise the use of hedonic imputation indexes when parameters are unstable, a proposal which will be considered in sections 5 and 7 below.

Section 2 below looks at a simple unweighted two period time dummy variable hedonic regression model. We focus on the estimation of the time dummy estimate of the change in log prices going from period 0 to 1 but we represent this measure of overall log price change as a difference in log price levels for the two periods. In section 3, we take the same unweighted model but run separate hedonic regressions for both periods and use these regression parameters to form two imputed measures of constant quality log price change. These two measures are then averaged to obtain an overall imputed measure of log price change.\textsuperscript{5} An exact expression for the difference in constant quality log price change between the time dummy and imputation measures is also developed in section 3. It is found that in order for these two overall measures to differ, we require:


\textsuperscript{4} The first exception is Silver and Heravi (2007a) who considered the case of one characteristic and used a rather different methodological approach based on the bias generated by omitted variables in regression models. The second exception is the comment by Jan de Haan which follows this paper, who developed an expression for the difference based on a framework outlined in Tripplett and MacDonald (1977). The points made in Haan’s commentary are developed more fully in Haan (2007).

\textsuperscript{5} An alternative interpretation of this measure of price change is derived in Appendix 1.
• differences in the two variance covariance matrices pertaining to the model characteristics in each period;
• differences in average amounts of model characteristics present in each period\(^6\) and
• differences in estimated hedonic coefficients for the two separate hedonic regressions.

The analysis in sections 2 and 3 is repeated in the weighted context in sections 4 and 5. Section 6 provides an empirical study for desktop PCs and section 7 concludes by discussing the issue of choice between the approaches in light of the theoretical and empirical findings.

Appendix 1 considers two alternative methodologies for constructing measures of overall log price change using the hedonic imputation methodology where two separate hedonic regressions are estimated for the two period under consideration. The first methodological approach is due to Court (1939; 108) where individual prices in each period are quality adjusted using their characteristics vectors and the characteristics prices obtained from one of the two hedonic regressions and then the resulting quality adjusted prices are compared across the two periods. Finally, the resulting two measures of quality adjusted overall log price change are averaged. In the second methodological approach to hedonic imputation indexes, due originally to Griliches (1967), the mean vector of characteristics that pertains to the models observed in period 0 is calculated and then the distance between the two hedonic regressions at this mean characteristics point is calculated, which generates a first measure of overall price change (the Laspeyres measure of log price change). The Paasche measure of overall log price change is calculated using the mean vector of characteristics that pertains to the models observed in period 1 and then the distance between the two hedonic regressions at this mean characteristics point is calculated. Finally the two estimates of overall log price change are averaged. Appendix 1 shows that these two methodological approaches to hedonic imputation indexes lead to exactly the same numerical estimates of overall price change.

It is often thought that a major advantage of the time dummy variable method for obtaining measures of overall log price change is that a standard error for the log price change is obtained. In Appendix 2, a method for obtaining approximate standard errors for the Laspeyres and Paasche hedonic imputation measures of log price change is derived.

### 2. Unweighted Time Dummy Hedonic Regressions

We begin by considering a simple unweighted two period time dummy variable hedonic regression model. We assume that there are \(N(t)\) observations on the prices, \(p_n^t\), of various models \(n\) in period \(t\) for \(t = 0,1\). Observation \(n\) in period \(t\) has a vector of \(K\) characteristics associated with it, say \([z_{n1}^t, z_{n2}^t, ..., z_{nK}^t]\) for \(t = 0,1\) and \(n = 1,2,...,N(t)\). The time dummy regression model has the following form:

\[
\ln p_n^t \equiv y_n^t = \alpha_t + \sum_{k=1}^{K} z_{nk}^t \gamma_k + \epsilon_n^t; \quad t = 0,1 ; n = 1,2,...,N(t)
\]

where the \(\epsilon_n^t\) are independently distributed normal variables with mean 0 and constant variance and \(\alpha_0, \alpha_1, \gamma_1, ..., \gamma_K\) are parameters to be estimated. The parameters \(\alpha_0\) and \(\alpha_1\) are measures of the average level of constant quality prices of the items in period 0 and 1 respectively and the \(\gamma_1, ..., \gamma_K\) are quality adjustment factors for the \(K\) characteristics; i.e., \(\gamma_k\) is the contribution to the log price of the product of adding an extra unit of characteristic \(k\). Note that we have parameterized the time dummy hedonic regression model in a slightly different way to the way it is usually done since we do not have an overall constant term in the regression plus a time dummy variable in period 1; instead

---

\(^6\) If the models are exactly the same in the two periods being considered, then this set of differences will be zero and the characteristics variance covariance matrices will also be identical. Hence the hedonic time dummy and hedonic imputation estimates of price change will be identical under these conditions. Thus the two methods will give rise to substantial conflicting estimates only in markets where there are many new models being introduced into the marketplace or many disappearing models (or both).
we have separate constant terms for each period. The overall measure of logarithmic price change going from period 0 to period 1 is \( \alpha_1 - \alpha_0 \).\(^7\) Let \( 1_0 \) and \( 0_0 \) be vectors of ones and zeros of dimension \( N(t) \), let \( y^0_0 \) and \( y^1_1 \) be the \( N(0) \) and \( N(1) \) dimensional vectors of period 0 and 1 logarithms of product prices, let \( \varepsilon^0 \) and \( \varepsilon^1 \) be the \( N(0) \) and \( N(1) \) dimensional vectors of period 0 and 1 stochastic disturbances and let \( Z^0 \) and \( Z^1 \) be matrices of the product characteristics in periods 0 and 1 respectively. Then the model defined by (1) can be written in matrix notation as follows:

\[
\begin{align*}
(2) \quad y^0 & = 1_0 \alpha_0 + 0_0 \alpha_1 + Z_0 \gamma + \varepsilon^0 ; \\
(3) \quad y^1 & = 0_1 \alpha_0 + 1_1 \alpha_1 + Z_1 \gamma + \varepsilon^1 .
\end{align*}
\]

Let \( \alpha^*_0 , \alpha^*_1 , \gamma^*_1 , \ldots , \gamma^*_K \) be the maximum likelihood or least squares estimators for the parameters that appear in (2) and (3). Then letting \( e^0 \) and \( e^1 \) be the vectors of least squares residuals for equations (2) and (3) respectively, the following equations will be satisfied by the parameter estimates and the data:

\[
\begin{align*}
(4) \quad y^0 & = 1_0 \alpha^*_0 + 0_0 \alpha^*_1 + Z_0 \gamma^* + e^0 ; \\
(5) \quad y^1 & = 0_1 \alpha^*_0 + 1_1 \alpha^*_1 + Z_1 \gamma^* + e^1 .
\end{align*}
\]

Let \( y = [y^{0T} , y^{1T}]^T \) and \( e = [e^{0T} , e^{1T}]^T \) and define \( \phi^* = [\alpha^*_0 , \alpha^*_1 , \gamma^*_1 , \ldots , \gamma^*_K]^T \). Now rewrite (4) and (5) as

\[
(6) \quad y = X\phi^* + e.
\]

It is well known that the columns of the \( X \) matrix are orthogonal to the vector \( e \) of least squares residuals; i.e., we have

\[
(7) \quad X^Te = X^T[y - X\phi^*] = 0_{2+K}.
\]

The first two equations in (7) are equivalent to the following two equations:\(^8\)

\[
\begin{align*}
(8) \quad 1_0^Ty^0 & = N(0)\alpha^*_0 + 1_0^T Z_0 \gamma^* ; \\
(9) \quad 1_1^Ty^1 & = N(1)\alpha^*_1 + 1_1^T Z_1 \gamma^* .
\end{align*}
\]

Equations (8) and (9) can be used to solve for the following period 0 and 1 constant quality log price levels:

\[
\begin{align*}
(10) \quad \alpha^*_0 & = \{1_0^T y^0 / N(0)\} - \{1_0^T Z_0 \gamma^* / N(0)\} = 1_0^T [y^0 - Z_0 \gamma^*] / N(0) ; \\
(11) \quad \alpha^*_1 & = \{1_1^T y^1 / N(1)\} - \{1_1^T Z_1 \gamma^* / N(1)\} = 1_1^T [y^1 - Z_1 \gamma^*] / N(1) .
\end{align*}
\]

Note that \( 1_t^Ty^t / N(t) \) is the arithmetic average of the log prices in period \( t \) for \( t = 0,1 \). Furthermore, note that \( 1_0^T Z_0 / N(0) \) is the arithmetic average of the amounts of each characteristic that are present in the period 0 models and \( 1_1^T Z_1 / N(1) \) is the corresponding arithmetic average amount of each characteristic that is present in the period 1 models. Thus each \( \alpha_t^* \) is equal to the average of the log prices for the models present in period \( t \) less a quality adjustment consisting of the inner product of the characteristic prices \( \gamma^* \) with the average amount of each characteristic across the models that are present in period \( t \). Alternatively, the second set of equalities in equations (10) and (11) shows that each \( \alpha_t^* \) is equal to the arithmetic average of the quality adjusted log prices, \( y^t - Z^t \gamma^* \), for the models

---

\(^7\) Our method of parameterization is equivalent to the standard method for parameterizing a time dummy hedonic regression model which is to have a common constant term for the two periods and a time dummy variable for the second period regression.

\(^8\) Diewert (2003a; 335) (2003b; 39) and Silver and Heravi (2005) used this orthogonality method of proof to provide an interpretation of the hedonic time dummy in terms of quality adjusted prices.
present in that period. In any case, the (unweighted) hedonic time dummy estimate of the change in log prices going from period 0 to 1, $\text{LP}_{\text{HD}}$, is the following difference in the log price levels:

$$\text{(12)} \quad \text{LP}_{\text{HD}} = \alpha_1^* - \alpha_0^*.$$  

For later reference, we work out an expression for the estimated characteristic prices, $\gamma^*$. Recall equation (6), $y = X\gamma + e$, which defined the $\text{N(0)+N(1)}$ by 2+K matrix $X$. We rewrite $X$ as follows:

$$\text{(13)} \quad X = [V,Z]$$

where $Z^T = [Z_{0T}^T, Z_{1T}^T]$ and $V$ is an $\text{N(0)+N(1)}$ by 2 matrix which has the first column equal to $[1_0^T, 0_1^T]^T$ and second column equal to $[0_0^T, 1_1^T]^T$. Now solve the least squares minimization problem that corresponds to (6) in two stages. In the first stage, we condition on $\gamma$ and minimize with respect to the components of $\alpha = [\alpha_0, \alpha_1]^T$. The resulting conditional least squares estimator for $\alpha$ is:

$$\text{(14)} \quad \alpha(\gamma) = (V^TV)^{-1}V^T[y - Z\gamma].$$

The second stage minimization problem is the problem of minimizing $f(\gamma)$ with respect to the components of $\gamma$ where $f$ is defined as follows:

$$\text{(15)} \quad f(\gamma) = [y - Z\gamma - V\alpha(\gamma)]^T[y - Z\gamma - V\alpha(\gamma)]$$

$$= [y - Z\gamma - V(V^TV)^{-1}V^T(y - Z\gamma)]^T[y - Z\gamma - V(V^TV)^{-1}V^T(y - Z\gamma)]$$  \hspace{1em} \text{using (14)}

$$= [My - MZ\gamma]^T[My - MZ\gamma]$$

$$= [y - Z\gamma]^TM^TM[y - Z\gamma]$$

$$= [y - Z\gamma]^TM[y - Z\gamma]$$  \hspace{1em} \text{since } M = M^T 	ext{ and } M^2 = M

where the projection matrix $M$ is defined as follows:

$$\text{(16)} \quad M = I - V(V^TV)^{-1}V^T.$$  

A simple way to solve the problem of minimizing $f(\gamma)$ with respect to $\gamma$ is to make use of the third equality in (15); i.e., define the projections of $y$ and $Z$ onto $M$ as follows:

$$\text{(17)} \quad y^* = My; Z^* = MZ.$$  

Using definitions (17), it can be seen that

$$\text{(18)} \quad f(\gamma) = [y^* - Z^*\gamma]^T[y^* - Z^*\gamma].$$

Thus the solution to the second stage least squares minimization problem is:

$$\text{(19)} \quad \gamma^* = (Z^*T Z^*)^{-1}Z^*T y^*.$$  

Once $\gamma^*$ has been determined by (19), then we can use (14) or (8) and (9) to determine the least squares estimators for $\alpha_0^*$ and $\alpha_1^*$.

---

9 This methodology can be traced back to Court (1939; 109-111) as his hedonic suggestion number two. Note also that if the models are the same in the two periods being considered, then $\text{N(0)}$ equals $\text{N(1)}$ (equals N say) and $Z^0$ equals $Z^1$, so that the two characteristics matrices are identical and thus $\alpha_1^* - \alpha_0^* = 1_T^T[y^1 - Z^1\gamma^1]/N(1) - 1_0^T[y^0 - Z^0\gamma^0]/N(0) = 1_T^T y^1/N - 1_0^T y^0/N$, which is the arithmetic mean of the period 1 log prices less the arithmetic mean of the period 0 log prices. Thus under these conditions, there is no need to run a hedonic regression; the usual matched model methodology can be used.
Using the definition of $V$, it can be shown that the projection matrix $M$ defined by (16) is block diagonal, with the two main diagonal blocks $M^0$ and $M^1$ defined as follows:

\[(20) \quad M^0 = I_0 - I_0 I_1^T/N(0) ; M^1 = I_1 - I_1 I_1^T/N(1) \]

where $I_0$ and $I_1$ are identity matrices of dimension $N(0)$ and $N(1)$ respectively. Using (20), we can determine more precisely what the vector $y^*$ equal to $My$ and the matrix $Z^*$ equal to $MZ$ look like. Let $y^* = [y_0^*, y_1^*]$ and $Z^* = [Z_0^*, Z_1^*]$. Then using (20), we have:

\[(21) \quad y^*_t = y_t - 1_0 I_0^{-1} y_t^T y_t/N(t) ; t = 0,1 ; \]

\[(22) \quad Z^*_t = Z_t - 1_1 I_1^{-1} Z_t^T Z_t/N(t) ; t = 0,1. \]

Thus each projected vector $y^*_t$ is equal to the corresponding period $t$ log price vector $y_t$ less a vector of ones times the average of the log prices for period $t$, $\sum_{n=1}^{N(t)} y_n^t/N(t)$, and each projected matrix $Z^*_t$ is equal to the corresponding period $t$ characteristics matrix $Z_t$ less a column vector of ones times a row vector equal to the average of the characteristics in each model for period $t$, $[\sum_{n=1}^{N(t)} z_{n1}^t/N(t), \sum_{n=1}^{N(t)} z_{n2}^t/N(t), ..., \sum_{n=1}^{N(t)} z_{nK}^t/N(t)]$. Thus $y^*$ and $Z^*$ are simply the corresponding $y$ and $Z$ with the period means subtracted from each component.

Using the block diagonal structure of $M$, it can be verified that we have the following alternative representation for the least squares characteristics prices $\gamma^*$ defined by (19):

\[(23) \quad \gamma^* = (Z_0^*T Z_0^* + Z_1^*T Z_1^*)^{-1} [Z_0^*T y_0^* + Z_1^*T y_1^*]. \]

We now turn our attention to hedonic imputation indexes.

### 3. Unweighted Hedonic Imputation Indexes

Instead of running one hedonic regression where the same characteristics prices are used to quality adjust prices in each period, we can run two entirely separate hedonic regressions with separate characteristics prices, $\gamma^0$ in period 0 and $\gamma^1$ in period 1. Thus using the same notation as in section 2, our models now are:

\[(24) \quad y_0 = 1_0 \beta_0 + Z_0 \gamma^0 + \eta_0^0 ; \]

\[(25) \quad y_1 = 1_1 \beta_1 + Z_1 \gamma^1 + \eta_1^1 , \]

where $\eta_0^0$ and $\eta_1^1$ are independently distributed normal random variables with means zero and constant variance within each period. Let $\beta_0^*, \gamma_0^0, ..., \gamma_k^0$ be the maximum likelihood or least squares estimators for the parameters that appear in (24) and let $\beta_1^*, \gamma_1^1, ..., \gamma_k^1$ be the maximum likelihood or least squares estimators for the parameters that appear in (25). Then letting $u_0^*$ and $u_1^*$ be the vectors of least squares residuals for equations (24) and (25) respectively, the following equations will be satisfied by the parameter estimates and the data:

\[(26) \quad y_0^* = 1_0 \beta_0^* + Z_0 \gamma_0^* + u_0^0 ; \]

\[(27) \quad y_1^* = 1_1 \beta_1^* + Z_1 \gamma_1^* + u_1^1 . \]

The counterparts to equations (8) and (9) in the present context are:

\[(28) \quad I_0^T y_0 = N(0) \beta_0^* + 1_0^T Z_0 \gamma_0^* ; \]

\[(29) \quad I_1^T y_1 = N(1) \beta_1^* + 1_1^T Z_1 \gamma_1^* . \]

Equations (28) and (29) lead to the following counterparts to equations (8) and (9):
Recall that the hedonic time dummy estimate of the change in log prices going from period 0 to 1, $LP_{PD}$, was defined by (12) as the difference in the log price levels, $\alpha_1^* - \alpha_0^*$. In the present context, we cannot simply take the difference between $\beta_1^*$ and $\beta_0^*$ as a measure of constant quality log price change between periods 0 and 1, because the quality adjustment parameters, $\gamma_0^{*0}$ and $\gamma_1^{*1}$, are different between the two periods. However, we can use the period 0 parameters, $\gamma_0^{*0}$, to form estimates of quality adjusted log prices for the models present in period 1 and then take the average of the resulting quality adjusted log prices, which we denote by $\delta_1^* :$

$$
(32) \delta_1^* = \{1_1^T y^1/N(1)\} - \{1_1^T Z^1 \gamma_0^{*0}/N(1)\} = 1_1^T [y^1 - Z^1 \gamma_0^{*0}]/N(1).
$$

Note that the above estimate of a period 1 log price level is analogous to $\beta_1^*$ defined by (31) except that the period 0 hedonic quality adjustment factors, $\gamma_0^{*0}$, are used in (32) whereas the period 1 hedonic quality adjustment factors, $\gamma_1^{*1}$, were used in (31). Since the period 0 and 1 estimated price levels, $\beta_0^*$ and $\delta_1^*$, use the same quality adjustment factors $\gamma_0^{*0}$ in order to form constant quality log prices in each period, we can take the difference $\delta_1^* - \beta_0^*$ as a measure of quality adjusted log price change between periods 0 and 1. 10 We call this hedonic imputation measure of log price change a Laspeyres type measure of price change and denote it by $\phi_1^* = \delta_1^* - \beta_0^*$. This measure of overall log price change depends asymmetrically on the characteristics price vector $\gamma_0^{*0}$ that was obtained from the period 0 hedonic regression. It can be seen that we can obtain an alternative measure of log price change between the periods using the period 1 hedonic regression characteristics price vector $\gamma_1^{*1}$. Thus use the period 1 characteristics price vector, $\gamma_1^{*1}$, to form estimates of quality adjusted log prices for the models present in period 0 and then take the average of the resulting quality adjusted log prices, which we denote by $\delta_0^* :$

$$
(33) \delta_0^* = \{1_0^T y^0/N(0)\} - \{1_0^T Z^0 \gamma_1^{*1}/N(0)\} = 1_0^T [y^0 - Z^0 \gamma_1^{*1}]/N(0).
$$

Note that the above estimate of a period 0 log price level is analogous to $\beta_0^*$ defined by (30) except that the period 1 hedonic quality adjustment factors, $\gamma_1^{*1}$, are used in (33) whereas the period 0 hedonic quality adjustment factors, $\gamma_0^{*0}$, were used in (30). Since the period 0 and 1 estimated price levels, $\delta_0^*$ and $\beta_1^*$ use the same quality adjustment factors $\gamma_1^{*1}$ in order to form constant quality log prices in each period, we can take the difference $\beta_1^* - \delta_0^*$ as a second measure of log price change between periods 0 and 1. We call this hedonic imputation measure of log price change a Paasche type measure of price change and denote it by $\phi_0^* = \beta_1^* - \delta_0^*$. 11

Following Griliches (1971b; 7) and Diewert (2003b; 12), it seems preferable to take a symmetric average of the above two estimates of log price change over the two periods. We choose the arithmetic mean 12 as our symmetric average and define the (unweighted) hedonic imputation estimate of the change in log prices going from period 0 to 1, $LP_{HI}$, as follows: 13

$$
\phi^* = \frac{\phi_1^* + \phi_0^*}{2} = \frac{\delta_1^* - \beta_0^* + \beta_1^* - \delta_0^*}{2}.
$$

10 This basic idea can be traced 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).

11 In Appendix 2, we develop a simple method for obtaining approximate standard errors for the hedonic imputation Laspeyres and Paasche measures of log price change between the two periods.

12 If we chose to measure price change instead of log price change, then the arithmetic mean estimator of log price change converts into a geometric mean of the two measures of level price change, $\exp[\delta_1^* - \beta_0^*]$ and $\exp[\beta_1^* - \delta_0^*]$.

13 In Appendix 1, we show that the measure of quality adjusted change in log prices defined by (34), which followed the methodology due originally to Court (1939; 108), can also be interpreted as a measure of the distance between the two hedonic regressions. The principles of such measures were discussed in Griliches (1967) and Dhrymes (1971; 111-112) and further developed by Feenstra (1995) and Diewert (2003a; 341-344). Empirical studies using this approach include.
Condition (37) can be refined. Recall (23) in the previous section, which provided a formula for the hedonic time dummy vector of quality adjustment factors, $\gamma^*$. The techniques that were used to
establish (23) can be used in order to establish the following expressions for the period 0 and 1 least squares estimates $\gamma^0$ and $\gamma^1$ that appear in (26) and (27):

\[(38) \gamma^0 = (Z^{0\prime}Z^0)^{-1}Z^{0\prime}y^0; \quad \gamma^1 = (Z^{1\prime}Z^1)^{-1}Z^{1\prime}y^1\]

where the $y^0$ and $Z^0$ are the demeaned $y$ and $Z$ as in the previous section.\(^{14}\) Now premultiply both sides of (23) by the matrix $Z^{0\prime}Z^0 + Z^{1\prime}Z^1$ and we obtain the following equation:\(^{15}\)

\[(39) [Z^{0\prime}Z^0 + Z^{1\prime}Z^1]y = Z^{0\prime}y^0 + Z^{1\prime}y^1\]

\[= Z^{0\prime}Z^0[\gamma^0] + Z^{1\prime}Z^1[\gamma^1] + Z^{0\prime}Z^0[\gamma^0] + Z^{1\prime}Z^1[\gamma^1]\]

\[= Z^{0\prime}Z^0[\gamma^0] + Z^{1\prime}Z^1[\gamma^1]\]

Using (38), we can obtain a more "fundamental" condition in terms of variance covariance matrices for the equality of the variance covariance matrix.

Equation (39) tells us that if $\gamma^0$ equals $\gamma^1$, then $\gamma^*$ is necessarily equal to this common vector. We now use equation (39) in order to evaluate the following expression:

\[(40) 2[Z^{0\prime}Z^0 + Z^{1\prime}Z^1][\gamma^0] - 2Z^{0\prime}Z^0[\gamma^0] + Z^{1\prime}Z^1[\gamma^1]\]

\[= -[Z^{0\prime}Z^0][\gamma^0] - [Z^{1\prime}Z^1][\gamma^1]
\]

Now premultiply both sides of (40) by $1/2[Z^{0\prime}Z^0 + Z^{1\prime}Z^1]^{-1}$ and substitute the resulting expression for $1/2[\gamma^0 + \gamma^1] - \gamma$ into equation (35) in order to obtain the following expression for the difference between the hedonic dummy estimate of constant quality price change and the corresponding symmetric hedonic imputation estimate:

\[(41) LP_{HD} - LP_{HI} = -(1/2)[1_{T}Z^1/N(1)] - [1_{0}Z^0/N(0)][Z^{0\prime}Z^0 + Z^{1\prime}Z^1][\gamma^0 - \gamma^1].\]

Using (41), it can be seen that the hedonic time dummy and hedonic imputation measures of log price change will be identical if any of the following three conditions are satisfied:

- $1_{T}Z^1/N(1)$ equals $1_{0}Z^0/N(0)$ so that the average amount of each characteristic across models in each period stays the same or
- $Z^{1\prime}Z^1$ equals $Z^{0\prime}Z^0$ so that the model characteristics total variance covariance matrix is the same across periods\(^{16}\) or
- $\gamma^*$ equals $\gamma^0$ so that separate (unweighted) hedonic regressions in each period give rise to the same characteristics quality adjustment factors.\(^{17}\)

In the following sections, we will adapt the above material to cover the case of weighted hedonic regressions.

### 4. Weighted Time Dummy Hedonic Regressions

\(^{14}\) Note that $Z^{\prime}Z^0/N(t)$ can be interpreted as a vector of sample covariances between the log prices in period $t$ and the amounts of the characteristics present in the period $t$ models while $Z^{\prime}Z^0/N(t)$ can be interpreted as a sample variance covariance matrix for the model characteristics in period $t$. In the main text, we refer to $Z^{\prime}Z^0$ as a characteristics "total" variance covariance matrix.

\(^{15}\) In the single characteristic case, (40) tells us that $\gamma^*$ is a weighted average of $\gamma^0$ and $\gamma^1$.

\(^{16}\) This condition for equality is also somewhat anticipated.

\(^{17}\) Using (38), we can obtain a more “fundamental” condition in terms of variance covariance matrices for the equality of $\gamma^0$ to $\gamma^1$; namely the equality of $(Z^{0\prime}Z^0)^{-1}Z^{0\prime}y^0$ to $(Z^{1\prime}Z^1)^{-1}Z^{1\prime}y^1$.\)
We now consider a weighted two period time dummy variable hedonic regression model. We again assume that there are N(t) observations on the prices, \( p_{nt}^{i} \), of various models n in period t for t = 0,1 but we now assume that the quantities purchased for each model n in period t, \( q_{nt}^{i} \), are also observable. Model n in period t again has the vector of K characteristics associated with it, \([z_{n1}^{i}, z_{n2}^{i}, ..., z_{nk}^{i}]\) for t = 0,1 and n = 1,2,...,N(t). The expenditure share of model n in period t is

\[
(42) \ s_{nt}^{i} = p_{nt}^{i} q_{nt}^{i} / \sum_{i=1}^{N(t)} p_{nt}^{i} q_{nt}^{i}; \quad t = 0,1 \ ; \ n = 1,2,...,N(t).
\]

Let \( s^{i} = [s_{n1}^{i}, ..., s_{N(t)}^{i}]^{T} \) denote the period t expenditure share vector for t = 0,1 and let \( S^{i} \) denote the diagonal period t share matrix that has the elements of the period t expenditure share vector \( s^{i} \) running down the main diagonal for t = 0,1. The matrix \( (S^{i})^{1/2} \) is the square root matrix of \( S^{i} \); i.e., the positive square roots of the elements of the period t expenditure vector \( s^{i} \) are the diagonal elements of this diagonal matrix for t = 0,1. As in the previous sections, \( y^{i} \) is the vector of period t log price changes and \( Z^{i} \) is the period t model characteristics matrix for t = 0,1.

With the above notational preliminaries out of the way, the weighted time dummy regression model that is the counterpart to the unweighted model defined earlier by (2) and (3) is defined as follows:\(^{18}\)

\[
(43) \ (S_{0}^{i})^{1/2} y^{0} = (S_{0}^{i})^{1/2} [1_{0} \alpha_{0}^{*} + 0_{0} \alpha_{1}^{*} + Z^{0} y^{*}] + e^{0};
\]

\[
(44) \ (S_{1}^{i})^{1/2} y^{1} = (S_{1}^{i})^{1/2} [0_{1} \alpha_{0}^{*} + 1_{1} \alpha_{1}^{*} + Z^{1} y^{*}] + e^{1}
\]

where the \( e^{i} \) vectors have elements \( e_{nt}^{i} \) that are independently distributed normal variables with zero means and constant variances.

Let \( \alpha_{0}^{*}, \alpha_{1}^{*}, y_{1}^{*}, ..., y_{K}^{*} \) be the maximum likelihood or least squares estimators for the parameters that appear in (43) and (44). Then letting \( e_{0}^{i} \) and \( e_{1}^{i} \) be the vectors of least squares residuals for equations (43) and (44) respectively, the following equations will be satisfied by the parameter estimates and the data:

\[
(45) \ (S_{0}^{i})^{1/2} y^{0} = (S_{0}^{i})^{1/2} [1_{0} \alpha_{0}^{*} + 0_{0} \alpha_{1}^{*} + Z^{0} y^{*}] + e_{0}^{i};
\]

\[
(46) \ (S_{1}^{i})^{1/2} y^{1} = (S_{1}^{i})^{1/2} [0_{1} \alpha_{0}^{*} + 1_{1} \alpha_{1}^{*} + Z^{1} y^{*}] + e_{1}^{i}
\]

The counterparts to equations (10) and (11) are now:

\[
(47) \ s_{0}^{i} = s^{0T} y^{0} - s^{0T} Z^{0} y^{*} = s^{0T} [y^{0} - Z^{0} y^{*}];
\]

\[
(48) \ s_{1}' = s^{1T} y^{1} - s^{1T} Z^{1} y^{*} = s^{1T} [y^{1} - Z^{1} y^{*}]
\]

Note that \( s_{t}' y^{i} = \sum_{n=1}^{N(t)} s_{nt}^{i} \ln p_{nt}^{i} \) is the period t expenditure share weighted average of the log prices in period t for t = 0,1. Furthermore, note that \( s_{t}' Z^{i} \) is a 1 by K vector whose kth element is equal to the period t expenditure share weighted average \( \sum_{n=1}^{N(t)} s_{nt}^{i} z_{nk}^{i} \) of the amounts of characteristic k that are present in the period t models for t = 0,1 and k = 1,...,K. Thus each \( \alpha_{i}' \) is equal to the expenditure share weighted average of the log prices for the models present in period t less a quality adjustment consisting of the inner product of the characteristic prices \( y^{*} \) with an expenditure share weighted average amount of each characteristic across the models that are present in period t. Alternatively, the second set of equalities in equations (47) and (48) shows that each \( \alpha_{i}' \) is equal to the period t

\(^{18}\) Diewert (2003b; 26) explained the logic behind the weighting scheme in the regression model defined by (43) and (44). For additional material on weighting in hedonic regressions, see Diewert (2005; 563) (2006; 13). Basically, the form of weighting that is used in (43) and (44) leads to measures of price change that are comparable (in the case where the characteristics of the models can be defined by dummy variables) to the “best” measures of price change in bilateral index number theory. It should be noted that the present weighted model will be equivalent to the previous unweighted model (4) and (5) only if the number of observations in each period are equal, so that N(0) equals N(1), and each share component s_{nt}^{i} is equal to a common value.
expenditure share weighted average of the quality adjusted log prices, \( y^t - Z^t\gamma^* \), for the models present in that period. Now use (47) and (48) in order to define the following *weighted hedonic time dummy estimate of the change in log prices* going from period 0 to 1: \( \text{LP}_{\text{WHD}} \), is the following difference in the log price levels:

\[
(49) \quad \text{LP}_{\text{WHD}} = \alpha_1^* - \alpha_0^* = s^1_T[y^1 - Z^1\gamma^*] - s^0_T[y^0 - Z^0\gamma^*].
\]

Thus the weighted hedonic time dummy estimate of the change in log prices is equal to a period 1 expenditure share weighted average of the quality adjusted log prices, \( y^1 - Z^1\gamma^* \), less a period 0 expenditure share weighted average of the quality adjusted log prices, \( y^0 - Z^0\gamma^* \).

For later reference, we can adapt the methodology presented at the end of section 2 in order to work out an explicit expression for the estimated characteristic prices, \( \gamma^* \). Recall the definitions of the diagonal matrices \( S^0 \) and \( S^1 \). Define \( S \) as a block diagonal matrix that has the blocks \( S^0 \) and \( S^1 \) on the main diagonals and let \( S^{1/2} \) be the corresponding square root matrix; i.e., this matrix has the elements of the period 0 expenditure share \( s^0 \) and the period 1 expenditure share vector \( s^1 \) running down the main diagonal. Adapting the analysis in section 1, we need to replace the \( y \) vector in that section by \( S^{1/2}y \) and the \( Z \) matrix in that section by \( S^{1/2}Z \), the \( V \) matrix by \( S^{1/2}V \) and define the counterpart \( M^\circ \) to the projection matrix \( M \) defined by (16) as follows:

\[
(50) \quad M^\circ = I - S^{1/2}V(V^T S)\text{\(S^{1/2}\)}^{-1}V^T S^{1/2} = I - S^{1/2}VV^T S^{1/2}
\]

where the second equality in (50) follows from the fact that \( V^T S \) equals \( I_2 \), a two by two identity matrix. Now define \( y^\circ \) and \( Z^\circ \) in terms of \( M^\circ \) and the original \( y \) vector and \( Z \) matrix as follows:

\[
(51) \quad y^\circ \equiv M^\circ S^{1/2}y = [I - S^{1/2}VV^T S^{1/2}]S^{1/2}y = S^{1/2}[I - VV^T]y
\]

\[
(52) \quad Z^\circ \equiv M^\circ S^{1/2}Z = [I - S^{1/2}VV^T S^{1/2}]S^{1/2}Z = S^{1/2}[I - VV^T]Z
\]

The new vector of time dummy quality adjustment factors \( \gamma^* \) can now be defined as the following counterpart to (19):

\[
(53) \quad \gamma^* = (Z^\circ T Z^\circ)^{-1}Z^\circ T y^\circ.
\]

Once \( \gamma^* \) has been determined by (53), then we can use (47) and (48) to determine the weighted least squares estimators for \( \alpha_0^* \) and \( \alpha_1^* \).

It is possible to express the \( \gamma^* \) defined by (53) in a more transparent way using our definitions of the matrices \( Z \), \( V \) and \( S \). Recall that in section 2, the demeaned log price change vectors \( y^\circ \) defined by (20) and the demeaned characteristics matrices \( Z^\circ \) defined by (21) proved to be useful. In those definitions, we used simple unweighted means. In the present context, we use *expenditure share weighted average means* as follows:

\[
(54) \quad y^t = y^t - 1_{s^t}^T y^t; \quad \text{ \( t = 0,1 \) ;}
\]

\[
(55) \quad Z^t = Z^t - 1_{s^t}^T Z^t; \quad \text{ \( t = 0,1 \).}
\]
Let the nth component of the vector \( y^t \) be \( y_n^t \) and let the nth row of the \( N(t) \) dimensional matrix \( Z^t \) be \( z_n^t \) for \( t = 0,1 \) and \( n = 1,2,...,N(t) \). Then it can be shown that the vector of characteristics prices \( \gamma^t \) defined by (53) can be written in terms of the components of the demeaned log price change vectors \( y^t \), the components of the expenditure share vectors \( s^i \) and the rows of the demeaned characteristics matrices \( Z^t \) as follows:

\[
(56) \gamma^t = \left[ \sum_{n=1}^{N(t)} s_n^0 z_n^0^T + \sum_{n=1}^{N(t)} s_n^0 z_n^1^T z_n^1^T \right]^{-1} \left[ \sum_{n=1}^{N(t)} s_n^0 z_n^0 y_n^0 + \sum_{n=1}^{N(t)} s_n^1 z_n^1 y_n^1 \right].
\]

Note that \( \sum_{n=1}^{N(t)} s_n^0 z_n^0 z_n^1^T \) can be interpreted as a period \( t \) expenditure share weighted sample variance covariance matrix \( \gamma^t \) for the model characteristics present in period \( t \) and \( \sum_{n=1}^{N(t)} s_n^1 z_n^0 y_n^1 \) can be interpreted as a period \( t \) expenditure share weighted sample covariance matrix between the period \( t \) log prices and the characteristics of the models present in period \( t \).

We now generalize the analysis on unweighted hedonic imputation indexes presented in section 3 to the weighted case.

5. Weighted Hedonic Imputation Indexes

Using the notation explained in the previous section, the two separate weighted hedonic regressions that are counterparts to the separate unweighted regressions (24) and (25) are now (57) and (58) below:

\[
(57) (S^0)^{1/2} y^0 = (S^0)^{1/2} [10\beta_0^* + Z^0 \gamma^0] + \eta^0,
\]

\[
(58) (S^1)^{1/2} y^1 = (S^1)^{1/2} [11\beta_1^* + Z^1 \gamma^1] + \eta^1,
\]

where \( \eta^0 \) and \( \eta^1 \) are independently distributed normal random variables with means zero and constant variance within each period. Let \( \beta_0^*, \gamma_1^0, ..., \gamma_K^0 \) be the maximum likelihood or least squares estimators for the parameters that appear in (57) and let \( \beta_1^*, \gamma_1^1, ..., \gamma_K^1 \) be the maximum likelihood or least squares estimators for the parameters that appear in (58). Then letting \( u^0 \) and \( u^1 \) be the vectors of least squares residuals for equations (57) and (58) respectively, the following equations will be satisfied by the parameter estimates and the data:

\[
(59) (S^0)^{1/2} y^0 = (S^0)^{1/2} [10\beta_0^* + Z^0 \gamma^0] + u^0,
\]

\[
(60) (S^1)^{1/2} y^1 = (S^1)^{1/2} [11\beta_1^* + Z^1 \gamma^1] + u^1.
\]

The weighted counterparts to equations (30) and (31) are:

\[
(61) \beta_0^* = s^0 y^0 - s^0 Z^0 \gamma^0 = s^0 y^0 - s^0 Z^0 \gamma^0;
\]

\[
(62) \beta_1^* = s^1 y^1 - s^1 Z^1 \gamma^1 = s^1 y^1 - s^1 Z^1 \gamma^1
\]

As in section 3, we cannot simply take the difference between \( \beta_1^* \) and \( \beta_0^* \) as a measure of constant quality log price change between periods 0 and 1, because the quality adjustment parameters, \( \gamma^0 \) and \( \gamma^1 \), are again different between the two periods. As in section 3, we can use the period 0 parameters, \( \gamma^0 \), to form estimates of quality adjusted log prices for the models present in period 1 and then take the period 1 weighted average of the resulting quality adjusted log prices, which we denote by \( \delta_1^* \):

\[
(63) \delta_1^* = s^1 y^1 - s^1 Z^1 \gamma^0 = s^1 y^1 - Z^1 \gamma^0
\]

\[\text{Note that for the present weighted model, } \sum_{n=1}^{N(t)} s_n^1 z_n^1 z_n^1^T \text{ is a true period } t \text{ expenditure share weighted sample variance covariance matrix, whereas for the earlier unweighted model, the counterpart to this sample variance covariance matrix was } Z^t Z^t^T, \text{ which was the period } t \text{ total variance covariance matrix, equal to } N(t) \text{ times the sample variance covariance matrix for the characteristics present in period } t \text{ models.}\]
Note that the above estimate of a period 1 log price level is analogous to $\beta_1^*$ defined by (62) except that the period 0 hedonic quality adjustment factors, $\gamma^{0*}$, are used in (63) whereas the period 1 hedonic quality adjustment factors, $\gamma^{1*}$, were used in (62). Since the period 0 and 1 estimated price levels, $\beta_0^*$ and $\delta_1^*$, use the same quality adjustment factors $\gamma^{0*}$ in order to form constant quality log prices in each period, we can again take the difference $\delta_1^*$ less $\beta_0^*$ as a measure of quality adjusted log price change between periods 0 and 1.\(^20\) This measure of overall log price change depends asymmetrically on the characteristics price vector $\gamma^{0*}$ that was obtained from the period 0 hedonic regression. It can be seen that we can obtain an alternative measure of log price change between the periods using the period 1 hedonic regression characteristics price vector $\gamma^{1*}$. Thus use the period 1 characteristics price vector, $\gamma^{1*}$, to form estimates of quality adjusted log prices for the models present in period 0 and then take the period 0 weighted average of the resulting quality adjusted log prices, which we denote by $\delta_0^*$:

\[ (64) \delta_0^* = s_0^T y^0 - s_0^T Z_0^0 \gamma^{1*} = s_0^T [y^0 - Z_0^0 \gamma^{1*}] . \]

Note that the above estimate of a period 0 log price level is analogous to $\beta_0^*$ defined by (61) except that the period 1 hedonic quality adjustment factors, $\gamma^{1*}$, are used in (64) whereas the period 0 hedonic quality adjustment factors, $\gamma^{0*}$, were used in (61). Since the period 0 and 1 estimated price levels, $\delta_0^*$ and $\beta_1^*$ use the same quality adjustment factors $\gamma^{1*}$ in order to form constant quality log prices in each period, we can take the difference $\beta_1^*$ less $\delta_0^*$ as a second measure of log price change between periods 0 and 1.\(^21\)

Again following Diewert (2003b; 20) and Haan (2003; 14) (2004), it seems preferable to take a symmetric average of the above two measures of log price change over the two periods. We again choose the arithmetic mean as our symmetric average and define the weighted hedonic imputation estimate of the change in log prices going from period 0 to 1, $LP_{WHI}$, as follows:\(^22\)

\[ (65) LP_{WHI} = \frac{1}{2}[\delta_1^* - \beta_0^*] + \frac{1}{2}[\beta_1^* - \delta_0^*]. \]

\[ = \frac{1}{2}[s_1^T [y^1 - Z_1^1 \gamma^{0*}] - s_0^T [y^0 - Z_0^0 \gamma^{0*}] + s_0^T [y^1 - Z_1^1 \gamma^{1*}] - s_0^T [y^0 - Z_0^0 \gamma^{1*}]] \]

\[ = s_0^T [y^1 - Z_0^1 \gamma^{0*} + \frac{1}{2} \gamma^{1*}] - s_0^T [y^0 - Z_0^0 \gamma^{1*}]. \]

Recall that in the weighted hedonic time dummy method for quality adjusting log prices $y^t$ for each period $t$, we used the characteristics quality adjustments defined by $Z^t \gamma^t$, where $\gamma^t$ was a constant across periods vector of quality adjustment factors. Looking at the right hand side of (65), it can be seen that the weighted hedonic imputation method for quality adjusting log prices in each period is similar but now the period $t$ vector of quality adjustments is $Z^t [\frac{1}{2} \gamma^{0*} + \frac{1}{2} \gamma^{1*}]$ instead of $Z^t \gamma^t$.

Using (49) and (65), we can form the following expression for the difference in the overall log price change using the two weighted methods for quality adjustment:

\[ (66) LP_{WHD} - LP_{WHI} = s_1^T [y^1 - Z_1^1 \gamma^{*}] - s_0^T [y^0 - Z_0^0 \gamma^{*}] - (s_1^T [y^1 - Z_1^{\gamma^{0*}}] - s_0^T [y^0 - Z_0^{\gamma^{0*}}]) \]

\(^{20}\) Haan (2003; 12) defined the exponential of this measure of log price change as the geometric Laspeyres hedonic imputation index of price change, except that for the matched models in both periods, he used actual prices rather than predicted prices.

\(^{21}\) Haan (2003; 13) defined the exponential of this measure of log price change as the geometric Paasche hedonic imputation index of price change, except that for the matched models in both periods, he used actual prices rather than predicted prices. Diewert (2003b; 13-14) considered similar hedonic imputation indexes except that he worked with ordinary Paasche and Laspeyres type indexes rather than geometric Paasche and Laspeyres type indexes.

\(^{22}\) The exponential of this measure of price change is approximately equal to Haan’s (2003; 14) geometric mean of his geometric Paasche and Laspeyres hedonic imputation indexes, which he regarded as an approximation to Törnqvist hedonic imputation index.
LP

and the corresponding symmetric hedonic imputation estimate using weighted hedonic regressions in expression for the difference between the hedonic dummy estimate of constant quality price change:

\[ \gamma^* = \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \]

Condition (67) says that the expenditure share weighted average amount of each characteristic for the models present in period 1 equals the corresponding expenditure share weighted average amount of each characteristic for the models present in period 0. Condition (68) says that the time dummy vector of quality adjustment factors, \( \gamma \), is equal to the arithmetic average of the two separate weighted hedonic regression estimates for the quality adjustment factors, \( \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \).

As in section 3, condition (68) can be strengthened. Recall (56) in the previous section, which provided a formula for the hedonic time dummy vector of quality adjustment factors, \( \gamma \). Using the same notation to that used in the previous section, we can establish the following expressions for the period 0 and 1 weighted least squares estimates \( \gamma^{0*} \) and \( \gamma^{1*} \) that appear in (59) and (60):

\begin{align*}
(69) \quad & \gamma^{0*} = \left[ \sum_{n=1}^{N(0)} s_n \begin{bmatrix} 0 & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} Z_n^{0*} \\ Y_n^{0*} \end{bmatrix} \right]^{-1} \left[ \sum_{n=1}^{N(0)} s_n \begin{bmatrix} 0 & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} Z_n^{0*} \\ Y_n^{0*} \end{bmatrix} \right] \gamma \\
(70) \quad & \gamma^{1*} = \left[ \sum_{n=1}^{N(1)} s_n \begin{bmatrix} 1 & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} Z_n^{1*} \\ Y_n^{1*} \end{bmatrix} \right]^{-1} \left[ \sum_{n=1}^{N(1)} s_n \begin{bmatrix} 1 & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} Z_n^{1*} \\ Y_n^{1*} \end{bmatrix} \right] \gamma 
\end{align*}

where the \( \gamma^{0*} \) and \( \gamma^{1*} \) are the demeaned \( \gamma \) and \( \gamma \) as in the previous section. Now premultiply both sides of (56) by the matrix \( \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} Z_n^{1*} \right] \) and we obtain the following equation which is the weighted counterpart to (40):

\[ \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \gamma^{0*} + \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \gamma^{1*} = \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \gamma 
\]

Equation (71) tells us if \( \gamma^{0*} \) equals \( \gamma^{1*} \), then \( \gamma \) is necessarily equal to this common vector. We now use equation (71) in order to evaluate the following expression:

\begin{align*}
(72) \quad & 2 \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \left[ \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \right] - \gamma \\
& = \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \left[ \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \right] - \gamma \\
& - 2 \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \gamma^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \gamma^{1*} \\
& = \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \left[ \gamma - \gamma^{0*} - \gamma^{1*} \right] \\
& = - \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \left[ \gamma - \gamma^{0*} - \gamma^{1*} \right]
\end{align*}

Now premultiply both sides of (72) by (1/2)\( \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \) and substitute the resulting expression for \( \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \) into equation (66) in order to obtain the following expression for the difference between the hedonic dummy estimate of constant quality price change and the corresponding symmetric hedonic imputation estimate using weighted hedonic regressions in both cases:

\[ \text{LP}_{\text{WHD}} - \text{LP}_{\text{WHI}} = \left[ s^T Z^1 - s^T Z^0 \right] \left[ \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \right] - \gamma \\
= - \frac{1}{2} \left[ s^T Z^1 - s^T Z^0 \right] \left[ \sum_{n=1}^{N(0)} s_n Z_n^{0*} + \sum_{n=1}^{N(1)} s_n Z_n^{1*} \right] \left[ \frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} \right] - \sum_{n=1}^{N(0)} s_n Z_n^{0*} \left[ \gamma^{1*} - \gamma^{0*} \right] - \sum_{n=1}^{N(1)} s_n Z_n^{1*} \left[ \gamma^{1*} - \gamma^{0*} \right]. \]
Using (73), it can be seen that the weighted hedonic time dummy and the weighted hedonic imputation measures of log price change will be identical if any of the following three conditions are satisfied:

- \( s^{1^T}Z^1 \) equals \( s^{0^T}Z^0 \) so that the period expenditure share weighted amount of each characteristic across models in each period stays the same or
- \( \sum_{n=1}^{N(1)} s_n Z_n \) equals \( \sum_{n=1}^{N(0)} s_n \) so that the expenditure share weighted model characteristics variance covariance matrix is the same in the two periods or
- \( \gamma^1 \) equals \( \gamma^0 \) so that separate (weighted) hedonic regressions in each period give rise to the same characteristics quality adjustment factors.

Which weighted method of quality adjustment is “best”? If either the weighted average amounts of each characteristic are much the same in the two periods being considered so that \( s^{1^T}Z^1 \) is close to \( s^{0^T}Z^0 \), or if the expenditure share weighted model characteristics variance covariance matrices are similar across periods, or if the separate weighted hedonic regression quality adjustment factors do not change much across the two periods, then it will not matter much which method is used, which is the new result that is demonstrated in this paper. If however, \( s^{1^T}Z^1 \) is not close to \( s^{0^T}Z^0 \), the expenditure share weighted model characteristics variance covariance matrices are different across periods, and the two separate weighted hedonic regressions (57) and (58) generate very different estimates for the quality adjustment factors, \( \gamma^0 \) and \( \gamma^1 \), then the method of quality adjustment could matter. If (57) and (58) are run together as a pooled regression model and an F test rejects the equality of \( \gamma^0 \) and \( \gamma^1 \), then it seems sensible to use the weighted hedonic imputation method which does not depend on having \( \gamma^0 \) equal to \( \gamma^1 \) as does the hedonic time dummy method. If the F test does not reject the equality of \( \gamma^0 \) and \( \gamma^1 \) and there are a large number of characteristics in the model, then valuable degrees of freedom will be saved if the weighted time dummy hedonic regression model is used. However, in this case, since \( \gamma^0 \) and \( \gamma^1 \) are necessarily close, it should not matter much which method is used. Thus it seems that the hedonic imputation methods probably give rise to “better” quality adjustments than dummy variable methods. We will revisit this discussion in section 7 below.

6. Empirical illustration: desktop personal computers (PCs)

The empirical study is of the measurement of changes in the quality-adjusted monthly prices of British desktop PCs in 1998. The data are monthly scanner data from the bar-code readers of PC retailers. The data amounted to 7,387 observations (a particular make and model of a PC sold in a given month in an either specialized or non-specialized PC store-type) representing a sales volume of 1.5 million models worth £1.57 billion. Table 1 shows that for the January to February price comparison there were 584 matched models available in both months for the price comparison. However, for the January to December price comparison only 161 matched models were available with 509 unmatched “old” models (available in January, but unmatched in December) and 436 unmatched “new” models (available in December but unavailable in January for matching). For product markets where there are a high proportion of unmatched models, Silver and Heravi (2005) demonstrated why matched model indexes suffer from sample selectivity bias and why hedonic indexes should be used instead.

The calculation of hedonic indexes requires the estimation of hedonic regression equations. To simplify the illustration we first include only a single explanatory variable in the hedonic (price) regressions, the speed in MHz. The regressions were run separately for each month for the hedonic imputation indexes, and over January and the current month, including a dummy variable for the

---

23 Bear in mind that some of the indexes estimated in this paper are also weighted by shares of sales values and that the fall off in the coverage of the matched sample by sales is even more dramatic: for the January to December comparison matched models made up only 71% of the January sales value and a mere 12% of the December sales value.
latter, for the hedonic dummy indexes. The estimated coefficients for speed in the hedonic regressions were statistically significant coefficients with the expected positive signs.\textsuperscript{24}

Table 2 columns (1) and (2) show falls in the \textit{unweighted} hedonic dummy and hedonic imputation indexes of 73.2 and 76.4 percent respectively, a difference of 3.2 percentage points. Columns (4)-(6) show the constituent elements of equation (41) that make up this difference: column (4) is the change in mean characteristics, \[1_11^{T}Z_1/N(1) − 1_0^{T}Z_0/N(0)\]—average speed increased by 127 MHz. over the year; column (5) is the change in the (total) variance-covariance characteristics matrices relative to their sum in the two periods, \[Z_0^{T}Z_0^{0} + Z_{1T}^{T}Z_{1}^{0} − [Z_{1T}^{T}Z_{1}^{0} − Z_{0T}^{T}Z_{0}^{0}]\]—with one characteristic in this illustration it is the relative change in the variance of speed, falling in some early months but increasing thereafter; and column (6) is the change in the characteristic parameter estimates, \([\gamma^*−\gamma^{0*}]\)—the estimated parameters decreased over time. The decomposition of the difference between the hedonic dummy and imputation indexes given in column (3) is exact as demonstrated in column (7)—the difference is equal to one-half of the product of columns (4), (5) and (6), following equation (41). We stress that the decomposition is based on the \textit{product} of these constituent parts. If either of these changes is zero, then there will be no difference between the indexes. While it is clear that there were large increases in the mean speed of PCs over the year, column (4), they did not materialize in substantial difference between the formulas, being tempered by the smaller changes in columns (5) and (6). The formulation also provides insights into the factors behind any difference in the results from these methods. For example, in September and December the changes in the estimated parameters were about the same, yet the difference between the hedonic dummy and imputation indexes in column (3) was higher in December than in September, driven by the larger change in mean characteristics in December.

The indexes considered in Table 2 were unweighted and thus unrepresentative if models differ in their popularity. Conventional index number theory requires that price changes should be weighted by relative expenditures shares and the same requirement should apply to hedonic indexes. In Sections 4 and 5 above \textit{weighted} hedonic dummy and hedonic imputation indexes were formulated and equation (73) provided a decomposition of the difference between them. The results for PCs for weighted hedonic indexes, their difference, and the constituent elements underlying the difference are given in Table 3. Again the difference depends on the product of three terms: the change in the expenditure share weighted mean of each characteristic, \([s^{1T}Z^{1} − s^{0T}Z^{0}]\); the change in the expenditure share weighted characteristics variance-covariance matrix, \(\Sigma_{n=1}^{N(1)}N(0)_{0}^{*}0_{0}^{T}Z_{0}^{0}\); the multiplication is of the change in average speed by the parameter estimate for speed. The result is invariant to the units of measurement used for speed since an accordingly lower coefficient would result if the units of measurement for speed were say doubled.

\textsuperscript{24} The F-statistics for the null hypothesis of coefficients being equal to zero averaged 34.2 for hedonic imputation indexes and 53.4 for hedonic dummy indexes, consistently rejecting the null at a 0.01% level and lower. The explanatory power of the estimated equations were naturally low for this specification with a single explanatory variable, especially since they did not include dummy variables on brand. Details of estimates from a fully specified model are available from the authors.

\textsuperscript{25} The multiplication is of the change in average speed by the parameter estimate for speed. The result is invariant to the units of measurement used for speed since an accordingly lower coefficient would result if the units of measurement for speed were say doubled.
compared with December, but for the weighted results it does not. This is because the much higher December change in characteristics, column (4), is largely offset by the much lower change in the relative variance in column (5).

The empirical example above was limited to a single price-determining quality characteristic variable for illustrative purposes. The decomposition for more than one characteristic is similar in principle to the case of a single explanatory variable, but the constituent items of equation (73) are matrices and it is the product of these matrices that is required to account for the difference between the formulas. The regression estimates are now based on three quality characteristics, speed in MHz., the hard disk capacity (CAP), and random access memory (RAM), both in MB. The estimated coefficients for the three variables in the hedonic regressions were statistically significant with the expected positive signs. The result of the decomposition of the difference between weighted hedonic dummy and weighted hedonic imputation indexes based on multiple characteristics for the January with December comparison only (for ease of exposition) is presented below in the matrix format of equation (73) and as Table 4.

The weighted hedonic dummy index in December compared with January, based on the extended specification, fell by 51.5 percent, compared with the fall in the weighted hedonic imputation index of 63.3, a sizable difference of 11.8 percentage points. From Tables 2 and 3 we saw that weighting matters. Here we identify the importance of a fuller specification of the hedonic regression used and its effect on the magnitude of the index change and the spread between the two estimates. The fuller specification has led to an increase in the spread between the two indexes from 7.5 to 11.8 percentage points. Prices are estimated to have fallen further using the extended variable set: a fall for the weighted hedonic dummy hedonic of 51.5 compared with 43.1 and for the hedonic imputation index of 63.3 compared with 50.6 percent. Additional explanatory variables in a quality adjusted hedonic regression based index are probably preferable. Yet in spite of this illustration, a fuller specification of the hedonic regression need not necessarily lead to an increase in the difference between the formulas. If any component of an additional variable—its change in relative dispersion, covariance, parameter, or mean value—is negligible, then it will have little effect on the difference. Indeed, the overall impact of additional variables in the product of matrices in equation (73) may take a different sign and reduce the discrepancy. However, in general, product development in high technology products such as PCs takes the form of increased product differentiation (dispersion of characteristics values), improvements in many product dimensions at the same time (which will generally change means and covariances), and decreasing characteristic production costs and marginal utilities, as consumers realign their preferences to the new standards (which will generally lead to changes in parameter estimates). Thus our expectation is that the HI and HD approaches to measuring price change may frequently give different estimates of overall price change in dynamic markets.

The four terms in equation (73) are reproduced below as matrices, and as Table 4, for the December with January comparison, the product of which is equal to the difference between the formulas, i.e.

\[
- \frac{1}{2} \begin{bmatrix} Z^1 - Z_0^0 \end{bmatrix} \begin{bmatrix} \sum_{n=1}^{N(0)} s_n z_n 0^0 T z_n^0 + \sum_{n=1}^{N(1)} s_n z_n 1^1 T z_n^1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{n=1}^{N(0)} N(1) s_n z_n 1^1 T z_n^1 \end{bmatrix} \begin{bmatrix} \sum_{n=1}^{N(0)} N(0) s_n z_n 0^0 T z_n^0 \end{bmatrix} \begin{bmatrix} \eta^1 - \gamma^0 \end{bmatrix}
\]

\[
= -\frac{1}{2} \begin{bmatrix} 6245.2 & 5422.8 & 637.2 \end{bmatrix} \begin{bmatrix} 637.2 & 3626.1 & 1131.0 \end{bmatrix} \times \begin{bmatrix} 5422.8 & 55658.7 & 3626.1 \end{bmatrix} = \begin{bmatrix} 139593, 180140, 14405 \end{bmatrix} \begin{bmatrix} 5422.8 & 55658.7 & 3626.1 \end{bmatrix} \times \begin{bmatrix} 637.2 & 3626.1 & 1131.0 \end{bmatrix}
\]

26 The adjusted-R² were 0.40 and 0.31 for the hedonic regression equations in January and November respectively. The specification could be extended to include many more variables including dummy variables for brands. The exposition here was simplified to illustrate the decomposition for more than one explanatory variable.

27 There are statistical caveats to this statement, i.e., it may not be useful to include additional explanatory variables if there are insufficient degrees of freedom or if multicollinearity between included and omitted variables is strong.
The values of the elements in the first row or column of the matrices are for the characteristic variable speed, the second, the CAP, and third, RAM. The vectors and matrices in turn are the change in the mean value of the characteristics, which were all positive; the sum of the inverted variance-covariance characteristics matrix over the two months (the second matrix in (73) has been inverted); the change in the variance-covariance characteristics matrix over the two months; and the change in the estimated parameters. Thus in the third matrix, for example the diagonal values of: 1,728.1, −3,672.3, and 167.3 are the change in the variances of the characteristics speed, CAP and RAM respectively. Note that in the last matrix, the parameter estimates for speed and RAM fall over the period by a similar amount, but the increase in the average value of speed (in the first matrix) is much more than that of the mean RAM size, and thus is a more significant driver of the difference between the formulas. There is also a marked increase in the average CAP, yet its parameter difference estimate at 0.00127 is positive, so that a higher marginal value is attached to it in December as compared with January, yet the absolute value of this change, and thus its impact on the difference, is lower than the corresponding change for speed.

7. Concluding Remarks on the Choice Between Hedonic Imputation Indexes and Hedonic Dummy Hedonic Indexes

Having identified the factors behind the difference between the hedonic dummy and imputation indexes, we turn to consider which if any formula is more appropriate. Given that both approaches make symmetric use of information in the two periods and can be formulated to have the same functional form and weighting system, a plausible stance, when they produce different results, is to take a (geometric) mean of the two. Yet there may be reasons to prefer one against the other.

The main concern with the use of the hedonic time dummy index approach, as given by the respective unweighted and weighted equations (12) and (49), is that by construction, it constrains the parameters on the characteristic variables to be the same. Berndt and Rappaport (2001) found, for example, from 1987 to 1999 for desktop PCs, the null hypothesis of adjacent-year equality of the characteristics parameters to be rejected in all but one case. For mobile PCs the null hypothesis of parameter stability was rejected in eight of the 12 adjacent-year comparisons. Berndt and Rappaport (2001) preferred the use of hedonic imputation indexes if there was evidence of parameter instability. Pakes (2003), using quarterly data for hedonic regressions for desktop PCs over the period 1995 to 1999, rejected just about any hypothesis on the constancy of the coefficients. He also advocated hedonic imputation indexes on the grounds that “...since hedonic coefficients vary across periods it [the hedonic dummy approach] has no theoretical justification.” Pakes (2003: 1593). The hedonic imputation method is inherently more flexible (in that it can deal with changes in purchasers’ valuations of characteristics over the periods being compared) than the hedonic dummy method and this is a big advantage for the HI method.

Note also that the difference between the two approaches has been found to depend on three change factors: the change in the mean characteristics, relative variance-covariance characteristics matrix, and parameter estimates. More specifically it was found that the difference depends on the product of such changes. As such, parameter instability by itself need not by itself be a cause for concern. Even if parameters were unstable, the difference between the indexes may be compounded or mitigated by a small change in any of the other components.
Nevertheless, the essence of the HD method is that only one regression is run, with the data in both periods appearing as dependent variables and with the restriction that the characteristics are valued at common “prices” for the two periods. In this interpretation, HD is not as flexible because of these restrictions. Why are these restrictions imposed? Perhaps for three reasons:

- To conserve degrees of freedom.
- To give an unambiguous estimate of the amount of price change going from period 0 to 1. Because the regression surfaces are parallel, we can measure the distance between the two surfaces at any characteristics point $z$ and get the same estimate of log price change, which is not the case in the HI methods.
- To minimize the influence of outliers, particularly in situations where degrees of freedom are small.

In view of the above considerations, the advantages and disadvantages of the two methods can be seen: HI is “better” because it allows for changing characteristics prices over time; i.e., it is more “flexible” but at the cost of:

- Using up more degrees of freedom and
- At the cost of leading to a less reproducible estimate of overall price change between the two periods, since we have to condition on one or more “reasonable” $z$ points to measure the distance between the two surfaces.

In practice, the last objection is not very serious; the Laspeyres and Paasche type estimates of price change are well established in index number theory as is the idea that these equally valid estimates of price change should be averaged in order to come up with a single measure of price change.

Thus all things considered, we favour HI methods unless degrees of freedom are very limited.

Triplett (2004) recognized that extensive product differentiation with a high model turnover is an increasing feature of product markets. The motivation for the use of hedonic regression techniques lies in the failure of the matched models method to adequately deal with price measurement in this context. Schultze and Mackie (2002) argued that hedonic indexes were the most promising approach to measuring price changes for such product markets, but advised that further research into such methods was needed: in particular, under what conditions will HD and HI measures of price change be different? This paper has provided answers to this question.

**Appendix 1: An Alternative Interpretation of the Hedonic Imputation Estimate of Log Price Change**

In section 3 above, we derived an estimator for the logarithm of overall price change between the two periods, $L_{\text{HI}}$, defined by (34), which we called the hedonic imputation estimate of the change in log prices going from period 0 to 1. We derived this estimator of price change following the methodology pioneered by Court (1939; 108); i.e., individual prices in each period were quality adjusted using their characteristics vectors and the characteristics prices obtained from one of the two hedonic regressions pertaining to the two periods under consideration and then the resulting quality adjusted prices were compared across the two periods. However, there is an alternative method for estimating price change across two periods when separate hedonic regressions are run for each period. In this second method, we calculate the mean vector of characteristics that pertains to the models observed in period 0 say and then calculate the distance between the two hedonic regressions at this mean characteristics point. This is called a Laspeyres type measure of price change. Then we calculate the mean vector of characteristics that pertains to the models observed in period 1 and calculate the distance between the two hedonic regressions at this second mean characteristics point,
leading to a Paasche type measure of price change. Finally these two distances between the hedonic regression surfaces are averaged in order to obtain a final measure of price change between the two periods. This is the methodology originally proposed by Griliches (1967) and Dhrymes (1971; 111-112). In this Appendix, we show that the first and second hedonic imputation methods lead to the same overall estimate of price change.

We first consider the unweighted hedonic imputation model that was described in section 3. Recall the notation used in section 3 above where \( y_0 \) was the \( N(0) \) dimensional vector of log model prices in period 0, \( y_1 \) was the \( N(1) \) dimensional vector of log model prices in period 1 and \( Z_t \) was the \( N(t) \) by \( K \) matrix of characteristics by model in period \( t \) for \( t = 0,1 \). Define the sample average of the log prices in period \( t \), \( y_t^* \), and the sample average vectors of model characteristics, \( z_t^* \), as follows:

\[
(A1) \quad y_{0*}^* = \frac{1}{T} y_0^*; \quad y_{1*}^* = \frac{1}{T} y_1^*; \quad z_{0*}^* = \frac{1}{T} Z_0^*; \quad z_{1*}^* = \frac{1}{T} Z_1^* \.
\]

Using definitions (A1), equations (28) and (29) in section 3 can be rewritten as follows:

\[
(A2) \quad y_{0*}^* = \beta_{0*}^* + z_{0*}^* \gamma_{0*}^*; \\
(A3) \quad y_{1*}^* = \beta_{1*}^* + z_{1*}^* \gamma_{1*}^*.
\]

For later reference, it can be seen that equations (A2) and (A3) imply the following expression for the difference in intercepts in the two hedonic regressions:

\[
(A4) \quad \beta_{1*}^* - \beta_{0*}^* = y_{1*}^* - y_{0*}^* - z_{1*}^* \gamma_{1*}^* + z_{0*}^* \gamma_{0*}^*.
\]

Now we are ready to define some estimators of the distance between the two hedonic regression surfaces. We define the Laspeyres type measure of log price change between periods 0 and 1, \( L_{PL} \), and the Paasche type measure of log price change, \( L_{PP} \), as follows:

\[
(A5) \quad L_{PL} = \beta_{1*}^* + z_{0*}^* \gamma_{1*}^* - [\beta_{0*}^* + z_{0*}^* \gamma_{0*}^*]; \\
(A6) \quad L_{PP} = \beta_{1*}^* + z_{1*}^* \gamma_{1*}^* - [\beta_{0*}^* + z_{1*}^* \gamma_{0*}^*].
\]

It can be seen that (A5) and (A6) are both measures of the distance between the two hedonic regression surfaces: the Laspeyres type measure holds the characteristics vector constant at the average of the period 0 levels, \( z_{0*}^* \), while the Paasche type measure holds the characteristics vector constant at the average of the period 1 levels, \( z_{1*}^* \). Our final measure of log price change is the arithmetic average of the Laspeyres and Paasche type measures, which we call the Fisher type measure of log price change, \( L_{PF} \):

\[
(A7) \quad L_{PF} = (1/2)[L_{PL} + L_{PP}]
\]

\[
= (1/2)[\beta_{1*}^* + z_{0*}^* \gamma_{1*}^* - (\beta_{0*}^* + z_{0*}^* \gamma_{0*}^*) + \beta_{1*}^* + z_{1*}^* \gamma_{1*}^* - (\beta_{0*}^* + z_{1*}^* \gamma_{0*}^*)]
\]

\[
= \beta_{1*}^* - \beta_{0*}^* + (1/2)[z_{0*}^* + z_{1*}^*][\gamma_{1*}^* - \gamma_{0*}^*]
\]

\[
= y_{1*}^* - y_{0*}^* - z_{1*}^* \gamma_{1*}^* + z_{0*}^* \gamma_{0*}^* + (1/2)[z_{0*}^* \gamma_{1*}^* - z_{0*}^* \gamma_{0*}^* + z_{1*}^* \gamma_{1*}^* - z_{1*}^* \gamma_{0*}^*]
\]

\[
= y_{1*}^* - y_{0*}^* + (1/2)[z_{0*}^* \gamma_{1*}^* + z_{0*}^* \gamma_{0*}^* - z_{1*}^* \gamma_{1*}^* - z_{1*}^* \gamma_{0*}^*] \\
\]

\[
= y_{1*}^* - y_{0*}^* - (1/2)[\gamma_{1*}^* + \gamma_{0*}^*][z_{1*}^* - z_{0*}^*] \\
\]

\[
= L_{PH}
\]

where \( L_{PH} \) was the hedonic imputation index defined by (34) in section 2. Thus the two hedonic imputation methods for defining an estimate of price change coincide in the unweighted case.\(^{28}\)

\(^{28}\) It can also be verified that \( L_{PL} \) is equal to \( \beta_{1*}^* - \delta_{0*}^* \) where \( \delta_{0*}^* \) was defined by (33) and \( L_{PF} \) is equal to \( \delta_{1*}^* - \beta_{0*}^* \) where \( \delta_{1*}^* \) was defined by (32).
Now consider the weighted hedonic imputation model that was described in section 5. The equally
weighted sample averages of the log prices \((y_0^*\) and \(y_1^*)\) and of the model characteristics (the vectors \(z_0^*\) and \(z_1^*)\) defined in (A1) above are now replaced by the following expenditure share weighted averages:

\[
(A8) \quad y_0^* \equiv s_0^T y_0; \quad y_1^* \equiv s_1^T y_1; \quad z_0^* \equiv s_0^T Z_0; \quad z_1^* \equiv s_1^T Z_1. 
\]

Using the new definitions in (A8), it can be seen that equations (61) and (62) in section 5 imply that equations (A2) and (A3) continue to hold so that (A4) also holds using these new definitions.

We can again define the Laspeyres and Paasche type measures of log price change by (A5) and (A4) where we use the new hedonic regression estimates for the period 0 weighted regression, \(\beta_0^*\) and \(\gamma_0^*\), and for the period 1 weighted regression, \(\beta_1^*\) and \(\gamma_1^*\), and the period 0 weighted average characteristics vector \(z_0^*\) for the Laspeyres measure \(LP_L\) and the period 1 weighted average characteristics vector \(z_1^*\) for the Paasche measure \(LP_P\). Now use \(LP_L\) and \(LP_P\) to define the Fisher measure \(LP_F\) by the first line in (A7) and again we can show that this Fisher measure is equal to the weighted hedonic imputation index \(LP_{WHI}\) defined by (65).

Thus we have shown that two rather different looking approaches to hedonic imputation indexes are equivalent.

**Appendix 2: A Method for Obtaining Approximate Standard Errors for the Hedonic Imputation Laspeyres and Paasche Measures of Log Price Change**

We consider the unweighted case first. Recall that the Laspeyres type hedonic imputation measure of log price change was defined as

\[
(A9) \quad \phi_L^* = \delta_1^* - \beta_0^* 
\]

where \(\delta_1^*\) and \(\beta_0^*\) are defined by (32) and (30) respectively. These last two equations can be rewritten as follows:

\[
(A10) \quad N(1) \delta_1^* = 1_1^T y_1 - 1_1^T Z_1^\gamma_0^*; \\
(A11) \quad N(0) \beta_0^* = 1_0^T y_0 - 1_0^T Z_0^\gamma_0^*. 
\]

Recall also that the period 0 hedonic regression was written as (26) where \(\beta_0^*\) was the period 0 estimated log price level and \(\gamma_0^*\) was the period 0 vector of least squares estimates for the characteristics prices. Now use these estimated period 0 regression coefficients to quality adjust the period 1 log prices in the vector \(y_1\). After subtracting these quality adjustments from the vector of period 1 log prices \(y_1\), we are left with the period 1 vector \(v_1\) of quality adjusted prices less the period 0 log price level defined as follows:

\[
(A12) \quad v_1 \equiv y_1 - [1_1^T \beta_0^* + Z_1^\gamma_0^*] = [y_1 - Z_1^\gamma_0^*] - 1_1^T \beta_0^*. 
\]

Now run a least squares regression of the period 1 residual vector \(v_1\) on a constant term with coefficient \(\phi_0\). The resulting least squares estimator for \(\phi_0\) is:

\[
(A13) \quad \phi_0^* = 1_1^T v_1 / N(1) \\
= [1_1^T [y_1 - Z_1^\gamma_0^*] / N(1)] - N(1) \beta_0^* / N(1) \quad \text{using (A12)} \\
= \delta_1^* - \beta_0^* \quad \text{using (A10)} \\
= \phi_L^* \quad \text{using (A9)}. 
\]
Thus the Laspeyres type hedonic imputation measure of log price change $\phi_L^*$ defined by (A9) is numerically equal to the least squares estimator $\phi_0^*$ of the constant term in a regression of period 1 quality adjusted log prices $v^1$ defined by (A12) on a constant and the standard error on this auxiliary regression coefficient can serve as an approximate standard error for the Laspeyres hedonic imputation measure of constant quality log price change over the two periods under consideration.  

The above algebra can be repeated for the Paasche type hedonic imputation measure of log price change, which was defined as

$$\text{(A14)} \quad \phi_P^* = \beta_1^* - \delta_0^*$$

where $\beta_1^*$ and $\delta_0^*$ are defined by (31) and (33) respectively. These last two equations can be rewritten as follows:

$$(\text{A15}) \quad N(1)\beta_1^* = 1_1^T y^1 - 1_1^T Z^1 \gamma^1.$$  

$$(\text{A16}) \quad N(0)\delta_0^* = 1_0^T y^0 - 1_0^T Z^0 \gamma^1.$$  

Recall also that the period 0 hedonic regression was written as (27) where $\beta_1^*$ was the period 1 estimated log price level and $\gamma^1$ was the period 1 vector of least squares estimates for the characteristics prices. Now use these estimated period 1 regression coefficients to quality adjust the period 0 log prices in the vector $y^0$. After subtracting these quality adjustments from the vector of period 0 log prices $y^0$, we are left with the period 0 vector $v^0$ of quality adjusted prices less the period 1 log price level defined as follows:

$$\text{(A17)} \quad v^0 = y^0 - [1_1^T \beta_1^* + Z^0 \gamma^1] = [y^0 - Z^0 \gamma^1] - 1_0^T \beta_1^*.$$  

Now run a least squares regression of the period 0 residual vector $v^0$ on a constant term with coefficient $\phi_1$. The resulting least squares estimator for $\phi_1$ is:

$$\text{(A18)} \quad \phi_1^* = 1_1^T v^0/N(0)$$

$$= {1_1^T [y^0 - Z^0 \gamma^1]}/N(0) - N(0)\beta_1^*/N(0) \quad \text{using (A17)}$$

$$= \delta_0^* - \beta_1^* \quad \text{using (A16)}$$

$$= -\phi_L^* \quad \text{using (A14)}.$$  

Thus the Paasche type hedonic imputation measure of log price change $\phi_P^*$ defined by (A14) is numerically equal to minus the least squares estimator $\phi_1^*$ of the constant term in a regression of period 0 quality adjusted log prices $v^0$ defined by (A17) on a constant and the standard error on this auxiliary regression coefficient can serve as an approximate standard error for the Paasche hedonic imputation measure of constant quality log price change over the two periods under consideration.  

We leave the reader with the task of deriving the counterparts of these results to the case where we have weighted hedonic regressions.

---

29 This is only an approximate standard error because it is conditional on period 0 estimated parameters, $\beta_0^*$ and $\gamma^0$, which are subject to some sampling uncertainty. If we wanted to use the same methodology to obtain standard errors for the constant quality period 0 log price level, $\beta_0$, and the constant quality period 1 log price level, $\delta_1$, then we would use the period 0 estimated characteristics prices $\gamma^0$ in order to form period 0 and 1 quality adjusted log price vectors $w^0$ and $w^1$ defined as $w^t = y^t - Z^t \gamma^0$ for $t = 0,1$. Now form two auxiliary regressions where $w^0$ is regressed on a constant with coefficient $\beta_0$ and $w^1$ is regressed on a constant with coefficient $\delta_1$. The least squares estimators for $\beta_0$ and $\delta_1$ turn out to be the $\beta_0^*$ and $\delta_0^*$ defined by (A10) and (A11). The standard errors for these coefficients in the auxiliary regression can be used as approximate standard errors for the log price levels in the two periods. These standard errors are conditional on the estimated period 0 characteristics prices, $\gamma^0$. Of course, the original period 0 hedonic regression can be used in order to obtain an unconditional standard error for the period 0 log price level $\beta_0$.  


References


<table>
<thead>
<tr>
<th>Month</th>
<th>Number of matched models</th>
<th>Number of unmatched old models in January of the comparison</th>
<th>Number of unmatched new models in the current month of the comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>February</td>
<td>584</td>
<td>86</td>
<td>104</td>
</tr>
<tr>
<td>March</td>
<td>577</td>
<td>93</td>
<td>181</td>
</tr>
<tr>
<td>April</td>
<td>346</td>
<td>324</td>
<td>191</td>
</tr>
<tr>
<td>May</td>
<td>315</td>
<td>355</td>
<td>227</td>
</tr>
<tr>
<td>June</td>
<td>297</td>
<td>373</td>
<td>265</td>
</tr>
<tr>
<td>July</td>
<td>282</td>
<td>388</td>
<td>301</td>
</tr>
<tr>
<td>August</td>
<td>276</td>
<td>394</td>
<td>351</td>
</tr>
<tr>
<td>September</td>
<td>247</td>
<td>423</td>
<td>382</td>
</tr>
<tr>
<td>October</td>
<td>193</td>
<td>477</td>
<td>402</td>
</tr>
<tr>
<td>November</td>
<td>164</td>
<td>506</td>
<td>435</td>
</tr>
<tr>
<td>December</td>
<td>161</td>
<td>509</td>
<td>436</td>
</tr>
</tbody>
</table>

Figures are for comparisons between January and each current month.
Table 2, Decomposition of differences between unweighted hedonic time dummy and imputation indexes for desktop PCs, 1998

<table>
<thead>
<tr>
<th></th>
<th>LP&lt;sub&gt;HD&lt;/sub&gt;</th>
<th>LP&lt;sub&gt;IM&lt;/sub&gt;</th>
<th>LP&lt;sub&gt;HD&lt;/sub&gt;-LP&lt;sub&gt;IM&lt;/sub&gt;</th>
<th>Change in mean characteristics</th>
<th>Relative change in var-cov matrices</th>
<th>Change in parameters</th>
<th>(4)×(5)×(6)/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change relative to fixed base, January 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>-0.0837</td>
<td>-0.0837</td>
<td>0.0000</td>
<td>-0.02</td>
<td>0.0564</td>
<td>-0.00042</td>
<td>0.0000</td>
</tr>
<tr>
<td>March</td>
<td>-0.1609</td>
<td>-0.1611</td>
<td>0.0003</td>
<td>7.69</td>
<td>0.1485</td>
<td>-0.00045</td>
<td>0.0003</td>
</tr>
<tr>
<td>April</td>
<td>-0.2717</td>
<td>-0.2713</td>
<td>-0.0004</td>
<td>47.16</td>
<td>-0.1895</td>
<td>-0.00010</td>
<td>-0.0004</td>
</tr>
<tr>
<td>May</td>
<td>-0.4334</td>
<td>-0.4348</td>
<td>0.0014</td>
<td>50.93</td>
<td>-0.1444</td>
<td>0.00039</td>
<td>0.0014</td>
</tr>
<tr>
<td>June</td>
<td>-0.4359</td>
<td>-0.4347</td>
<td>-0.0012</td>
<td>60.64</td>
<td>-0.0390</td>
<td>-0.00099</td>
<td>-0.0012</td>
</tr>
<tr>
<td>July</td>
<td>-0.4535</td>
<td>-0.4540</td>
<td>0.0004</td>
<td>68.99</td>
<td>0.0119</td>
<td>-0.00103</td>
<td>0.0004</td>
</tr>
<tr>
<td>August</td>
<td>-0.4560</td>
<td>-0.4672</td>
<td>0.0112</td>
<td>76.71</td>
<td>0.1484</td>
<td>-0.00196</td>
<td>0.0112</td>
</tr>
<tr>
<td>September</td>
<td>-0.4272</td>
<td>-0.4467</td>
<td>0.0195</td>
<td>84.15</td>
<td>0.1787</td>
<td>-0.00259</td>
<td>0.0195</td>
</tr>
<tr>
<td>October</td>
<td>-0.5727</td>
<td>-0.5945</td>
<td>0.0218</td>
<td>103.89</td>
<td>0.1906</td>
<td>-0.00220</td>
<td>0.0218</td>
</tr>
<tr>
<td>November</td>
<td>-0.6705</td>
<td>-0.7046</td>
<td>0.0341</td>
<td>118.27</td>
<td>0.2317</td>
<td>-0.00249</td>
<td>0.0341</td>
</tr>
<tr>
<td>December</td>
<td>-0.7320</td>
<td>-0.7643</td>
<td>0.0323</td>
<td>127.47</td>
<td>0.2050</td>
<td>-0.00247</td>
<td>0.0323</td>
</tr>
</tbody>
</table>
Table 3, Decomposition of differences between weighted hedonic time dummy and imputation indexes for desktop PCs, 1998

<table>
<thead>
<tr>
<th></th>
<th>LP_{HD} (1)</th>
<th>LP_{II} (2)</th>
<th>LP_{HD}-LP_{II} (3)</th>
<th>Change in mean characteristics (4)</th>
<th>Relative change in var-cov matrices (5)</th>
<th>Change in parameters (6)</th>
<th>(4)×(5)×(6)/2 (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change relative to fixed base, January 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>-0.0622</td>
<td>-0.0624</td>
<td>0.0002</td>
<td>8.84</td>
<td>0.0454</td>
<td>-0.00076</td>
<td>0.0002</td>
</tr>
<tr>
<td>March</td>
<td>-0.0920</td>
<td>-0.0921</td>
<td>0.0001</td>
<td>14.81</td>
<td>0.0122</td>
<td>-0.00092</td>
<td>0.0001</td>
</tr>
<tr>
<td>April</td>
<td>-0.1750</td>
<td>-0.1746</td>
<td>-0.0004</td>
<td>31.44</td>
<td>-0.0331</td>
<td>-0.00074</td>
<td>-0.0004</td>
</tr>
<tr>
<td>May</td>
<td>-0.2306</td>
<td>-0.2318</td>
<td>0.0012</td>
<td>42.74</td>
<td>0.0458</td>
<td>-0.00122</td>
<td>0.0012</td>
</tr>
<tr>
<td>June</td>
<td>-0.1873</td>
<td>-0.2002</td>
<td>0.0129</td>
<td>54.79</td>
<td>0.1464</td>
<td>-0.00321</td>
<td>0.0129</td>
</tr>
<tr>
<td>July</td>
<td>-0.2782</td>
<td>-0.2914</td>
<td>0.0131</td>
<td>65.76</td>
<td>0.1180</td>
<td>-0.00338</td>
<td>0.0131</td>
</tr>
<tr>
<td>August</td>
<td>-0.2974</td>
<td>-0.3271</td>
<td>0.0297</td>
<td>81.67</td>
<td>0.2299</td>
<td>-0.00316</td>
<td>0.0297</td>
</tr>
<tr>
<td>September</td>
<td>-0.2336</td>
<td>-0.3040</td>
<td>0.0704</td>
<td>94.45</td>
<td>0.3829</td>
<td>-0.00390</td>
<td>0.0704</td>
</tr>
<tr>
<td>October</td>
<td>-0.3891</td>
<td>-0.4476</td>
<td>0.0586</td>
<td>118.52</td>
<td>0.2979</td>
<td>-0.00332</td>
<td>0.0586</td>
</tr>
<tr>
<td>November</td>
<td>-0.5080</td>
<td>-0.5566</td>
<td>0.0486</td>
<td>131.90</td>
<td>0.2131</td>
<td>-0.00346</td>
<td>0.0486</td>
</tr>
<tr>
<td>December</td>
<td>-0.4311</td>
<td>-0.5057</td>
<td>0.0745</td>
<td>139.59</td>
<td>0.2766</td>
<td>-0.00386</td>
<td>0.0745</td>
</tr>
</tbody>
</table>
Table 4, Factors contributing to difference between weighted HD and weighted HI estimates for PCs for three variables: January to December comparison

<table>
<thead>
<tr>
<th>LP&lt;sub&gt;HI&lt;/sub&gt;</th>
<th>LP&lt;sub&gt;HI&lt;/sub&gt;</th>
<th>LP&lt;sub&gt;HD&lt;/sub&gt;-LP&lt;sub&gt;HD&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4311</td>
<td>-0.5057</td>
<td>0.0745</td>
</tr>
<tr>
<td>0.0001</td>
<td>-0.0004</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Factors contributing to change

<table>
<thead>
<tr>
<th>Change in mean characteristics</th>
<th>Change in variances</th>
<th>Change covariances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>139.593</td>
<td>Speed</td>
</tr>
<tr>
<td>CAP</td>
<td>180.140</td>
<td>CAP</td>
</tr>
<tr>
<td>RAM</td>
<td>14.405</td>
<td>RAM</td>
</tr>
</tbody>
</table>

Inverse of total variances

| Speed                         | 6,245.2             | Speed:CAP          |
| CAP                           | 55,658.7            | Speed:RAM          |
| RAM                           | 1,131.1             | CAP:RAM            |

Inverse of total covariances

| Speed:CAP                    | 5,422.8             | Speed              |
| Speed:RAM                    | 637.2               | CAP                |
| CAP:RAM                      | 3,626.1             | RAM                |

Change in parameter estimates

| Speed                         | -0.00743            |
| CAP                           | 0.00127             |
| RAM                           | -0.00612            |
Erwin Diewert, Saeed Heravi and Mick Silver: Hedonic Imputation versus Time Dummy Hedonic Indexes

Comment by Jan de Haan (Statistics Netherlands)

Hedonic regression has now become one of the standard tools for statistical agencies to adjust their CPIs for quality changes in markets with a high turnover of differentiated models such as PCs. The authors address an important question, namely the difference between ‘hedonic imputation indexes’ and time dummy hedonic indexes, which are the two main approaches to estimating hedonic price indexes (in the academic literature). They provide a novel exposition of the factors underlying the difference between these approaches, both for the unweighted and the preferred expenditure-share weighted case. In particular, the authors derive three conditions under which the two approaches lead to identical results: constancy (over time) of the average characteristics, constancy of the estimated characteristics parameters (used in the imputation approach), and constancy of the characteristics variance-covariance matrix. As the authors rightly claim, the third condition is somewhat unanticipated. Apart from being a valuable contribution to the literature, the paper seems highly relevant for the work of statistical agencies. However, the use of matrix algebra makes the exposition very technical, and the implications may not be readily understood by a typical price statistician (though the empirical illustration is certainly helpful). Below I present some of the authors’ findings in a simplified way by avoiding matrix notation, comment on them and take the opportunity to make a few additional observations. I focus on the unweighted case, just for the sake of simplicity, but spend a few words on weighting also.

A couple of choices have implicitly been made in the paper right from the start. For example, it is assumed that the hedonic regressions are run on the price data that are collected for the CPI. There may be statistical offices that perform hedonic regressions on a different data set and then use the estimated coefficients to adjust the raw CPI data for quality changes (which, I agree, is a problematic approach). More importantly, the paper discusses a specific type of hedonic imputation. Let $S^0$ and $S^1$ be the samples of items in periods 0 and 1; $S_M = S^0 \cap S^1$ is the matched sample with size $n_M$, $S_D$ the sub-sample of disappearing items, and $S_N$ the sub-sample of new items. For simplicity I assume a fixed sample size $n$; thus $n - n_M$ is the number of disappearing and new items. I
distinguish three types of unweighted symmetric imputation indexes: single imputation (SI), double imputation (DI) and full imputation (FI) indexes, as follows:

\[
\hat{P}_{SI} = \prod_{i \in S_1} \left( \frac{p_i^1}{\hat{p}_i^0} \right)^{\frac{1}{n}} \prod_{i \in S_0} \left( \frac{\hat{p}_i^0}{\hat{p}_i^0} \right)^{\frac{1}{2n}};
\]  

(1)

\[
\hat{P}_{DI} = \prod_{i \in S_1} \left( \frac{\hat{p}_i^1}{\hat{p}_i^0} \right)^{\frac{1}{n}} \prod_{i \in S_0} \left( \frac{\hat{p}_i^0}{\hat{p}_i^0} \right)^{\frac{1}{2n}};
\]  

(2)

\[
\hat{P}_{FI} = \prod_{i \in S_1} \left( \frac{\hat{p}_i^1}{\hat{p}_i^0} \right)^{\frac{1}{n}} \prod_{i \in S_0} \left( \frac{\hat{p}_i^0}{\hat{p}_i^0} \right)^{\frac{1}{2n}}.
\]  

(3)

where \( p_i^t \) denotes the price of item \( i \) in period \( t (t=0,1) \) and \( \hat{p}_i^t \) an imputed (predicted) price. If I am correct, the authors seem to consider (at least implicitly) a fourth type of imputation index, namely

\[
\hat{P}_{HI} = \left[ \prod_{i \in S_1} \left( \frac{\hat{p}_i^1}{p_i^0} \right)^{\frac{1}{n}} \prod_{i \in S_0} \left( \frac{\hat{p}_i^0}{\hat{p}_i^0} \right)^{\frac{1}{n}} \right]^\frac{1}{2n}
\]  

(4)

(in case of a fixed sample size). They use a log-linear hedonic model that explains the logarithm of price \( p_i^t \) from a set of \( K \) characteristics \( z_{ik} \) and an intercept term \( \alpha^t \):

\[
\ln(p_i^t) = \alpha^t + \sum_{k=1}^K \beta_k^t z_{ik} + \epsilon_i^t,
\]  

(5)

where \( \beta_k^t \) is the parameter for \( z_{ik} \). By assumption the random errors \( \epsilon_i^t \) have expected values of zero and constant (identical) variances.\(^\text{30}\) Equation (9) is estimated separately in periods 0 and 1, that is on the data of the samples \( S^0 \) and \( S^1 \). OLS regression of (5) yields parameter estimates \( \hat{\alpha}^t \) and \( \hat{\beta}_k^t \) and predicted prices \( \hat{p}_i^t = \exp[\hat{\alpha}^t + \sum_{k=1}^K \hat{\beta}_k^t z_{ik}] \). However, because the OLS regression residuals \( \epsilon_i^t = \ln(p_i^t) - \ln(\hat{p}_i^t) = \ln(p_i^t / \hat{p}_i^t) \) sum to zero in each period, i.e. \( \sum_{i \in S} \epsilon_i^t = 0 \), the index given by (4) coincides with the full imputation index (3). In Appendix 1 the authors show that this approach is equivalent to

\(^\text{30}\) Characteristics have no superscript for time \( t \) as an individual item is supposed to be of constant quality so that its characteristics are fixed over time.
what is often called the characteristics prices approach. The full imputation index (3), and thus (4), can be related to the single imputation index (1) in the following way:

\[
\hat{P}_{FI} = \hat{P}_{SI} \left[ \exp\left(\frac{\hat{e}_D^0}{\exp(\hat{e}_D^0)}\right) \right]^{1-f_M} \left[ \exp\left(\frac{\hat{e}_N^1}{\exp(\hat{e}_N^1)}\right) \right]^{1-f_M},
\]

where \( f_M = n_M / n \) denotes the fraction of matched items and \( 1 - f_M = (n - n_M) / n \) the fraction of unmatched items; \( \hat{e}_D^0 = \sum_{i \in S^0} e_i^0 / (n - n_M) \) and \( \hat{e}_N^1 = \sum_{i \in S^1} e_i^1 / (n - n_M) \) are the average residuals for the disappearing and new items. Dividing (2) by (1) yields

\[
\hat{P}_{DI} = \hat{P}_{SI} \left[ \exp(\hat{e}_D^0) / \exp(\hat{e}_N^1) \right]^{1-f_M},
\]

which relates the double imputation index (2) to the single imputation index. Equations (6) and (7) show that the choice of imputation method matters if the average residuals of the disappearing and new items differ, especially if they have different signs (and \( f_M \) is relatively small). For example, \( \hat{P}_{DI} < \hat{P}_{SI} < \hat{P}_{FI} \) if \( \hat{e}_D^0 < 0 \) and \( \hat{e}_N^1 > 0 \). This happens if disappearing items are sold at prices which are unusually low given their characteristics, perhaps due to ‘dumping’, and new items are introduced at unusually high prices.

At first sight it is not obvious why we would prefer the full imputation index (3), and thus (4), to the other imputation methods. A drawback seems to be that the observed prices are replaced by model-based estimates: in general this increases the variance of the hedonic index as it adds model variance to the matched-item part and may give rise to unnecessary bias if the hedonic model would be misspecified (which almost certainly happens to some extent in practice). But this view might be too simplistic. It can easily be shown that the following expression applies to the full imputation index:

\[
\hat{P}_{FI} = \prod_{t=1}^{T} \left[ \frac{1}{\prod_{t=1}^{T} \left[ \exp\left(\sum_{i \in S^0} \hat{\beta}_k^{0i} (z_k^0 - \bar{z}_k^0) \right) \right] \right],
\]

with \( \hat{\beta}_k^{0i} = (\hat{\beta}_k^0 + \hat{\beta}_k^1) / 2 \) and where \( z_k' = \sum_{i \in S} z_k / n \) is the average sample value of the \( k \)-th characteristic in period \( t \) \((t=0,1)\). By taking logs of (8) the authors’ equation (34) is obtained (for a fixed sample size). Since the average characteristics of matched items are the same across periods, by denoting the average characteristics of the disappearing and
new items by \[ z_{Dk} = \sum_{i \in S^D} z_{ik} / (n - n_M) \] and \[ z_{Nk} = \sum_{i \in S^N} z_{ik} / (n - n_M) \] respectively, expression (8) can be rewritten as

\[
\hat{P}_{it} = \left[ \prod_{i \in S^U} \left( \frac{p_i^1}{p_i^0} \right)^{1/n_M} \right]^{1 - f_M} \left[ \prod_{i \in S^D} \left( \frac{p_i^1}{p_i^0} \right)^{1/n_M} \right] \exp \left[ \sum_{k=1}^{K} \hat{\beta}_k^0 \left( z_{Dk}^0 - z_{Nk}^1 \right) \right] \right]^{1 - f_M},
\]

(9)

Equation (9) shows that this hedonic index is a weighted average of the matched-item index \[ \prod_{i \in S^U} \left( p_i^1 / p_i^0 \right)^{1/n_M} \] and a (quality-adjusted) index for the unmatched items. The latter index adjusts the ratio of geometric mean prices of new and disappearing items for differences in the average characteristics of those items. Equation (9) further shows that the matched items’ price relatives are implicitly left unchanged. Thus, matching where possible remains the basic principle even if a hedonic index would be estimated that, at first glance, does not seem to rely on matching.\(^{31}\)

Another advantage of the full imputation approach is its comparability with the time dummy approach. In its standard form the time dummy hedonic model reads

\[
\ln(p_i^t) = \alpha + \delta D_i^t + \sum_{k=1}^{K} \beta_k^0 z_{ik} + \epsilon_i^t,
\]

(10)

where \( D_i^t \) is a dummy variable that takes on the value of 1 if \( i \) is sold period \( t \) (i.e. for \( i \in S^t \)) and 0 otherwise (for \( i \in S^0 \)). A pooled OLS regression yields predicted prices \( \hat{p}_i^0 = \exp[\hat{\alpha} + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{ik}] \) and \( \hat{p}_i^1 = \exp[\hat{\alpha} + \hat{\delta} + \sum_{k=1}^{K} \hat{\beta}_k^1 z_{ik}] \). It follows that

\[
\hat{P}_{TD} = \exp(\hat{\delta}) = \frac{\hat{p}_i^1}{\hat{p}_i^0} = \frac{\prod_{i \in S^D} \left( p_i^1 / p_i^0 \right)^{1/n}}{\prod_{i \in S^D} \left( p_i^0 \right)^{1/n}} \exp \left[ \sum_{k=1}^{K} \hat{\beta}_k^0 \left( z_{Dk}^0 - z_{Nk}^1 \right) \right],
\]

(11)

using the fact that, since an intercept term is included in (10), the residuals again sum to zero in both periods. Expression (11) is well known (see for example Triplett, 2004) and

\(^{31}\)This is of course not to say that statistical agencies should try to match as much as possible. Samples preferably reflect the population of items at any point in time, possibly by using (or trying to mimic) PPS sampling. The fact that the current period sample may differ substantially from the base period sample is what makes the use of hedonics so important.
is quite similar to (8). This means that the time dummy index can also be written in the form of equation (9) when we replace $\hat{\beta}_k^{01}$ by $\hat{\beta}_k$. Using this result we obtain

$$\hat{P}_{TD} = \exp \left[ (1 - f_M) \sum_{k=1}^{K} (\hat{\beta}_k - \hat{\beta}_k^{01}) (z_{Dk}^0 - z_{Nk}^1) \right] \hat{P}_{FY},$$  \hspace{1cm} (12)$$

which makes clear that the difference between the time dummy index and the hedonic imputation will particularly be small if the set of matched items is large, the (average) regression coefficients from both approaches are close to each other, and the differences in the average characteristics of the new and disappearing items are small.

Finally I turn to weighted hedonic price indexes. The authors choose expenditure shares pertaining to the single period as regression weights. The advantage is obvious: it is a straightforward generalisation of the unweighted approach, yielding an estimator of the (full) imputation Törnqvist price index. The same set of regression weights is used for the weighted time dummy approach, so that both weighted approaches can easily be compared. The disadvantage, on the other hand, is that WLS regression might increase the variance of the estimated parameters compared to OLS (especially if, as the authors assume, the errors have identical variances). The use of the single imputation Törnqvist index

$$\hat{P}_{TI} = \prod_{i \in S_M} \left( \frac{p_i^1}{p_i^0} \right)^{\frac{\delta_i}{\mu_i}} \prod_{i \in S_D} \left( \frac{\hat{p}_i^1}{\hat{p}_i^0} \right)^{\frac{\delta_i}{2 \mu_i}} \prod_{i \in S_N} \left( \frac{p_i^1}{p_i^0} \right)^{\frac{\delta_i}{2}},$$  \hspace{1cm} (13)$$

or its double imputation counterpart, is however more flexible: explicit weighting makes it possible to apply all kinds of regression weights when estimating model (5), including equal weights. Moreover, the authors’ weighted time dummy index violates the (weak) identity test in a matched-item context (without new or disappearing items), a property that was mentioned already by Diewert (2003).\footnote{De Haan (2004), following up on Diewert (2003), proposed using regression weights for the WLS time dummy approach that are identical to the weights of the price relatives in equation (14). In that case the time dummy index can be interpreted as a single imputation Törnqvist index and of course in a matched-item situation the matched-item Törnqvist will be obtained.} My conclusion would be that the issue of weighting in hedonic regressions is still unresolved.
Response by Erwin Diewert, Saeed Heravi and Mick Silver to the Commentary by Jan de Haan

We are very grateful for the comments made by Jan de Haan on our paper. In particular, his equations (1)-(4) make clear the various alternatives that could be used by statistical agencies in constructing elementary price indexes using hedonic regressions to quality adjust new and disappearing items or models for a narrowly specified commodity. The commentary by Haan provides statistical agencies with a very useful overview of the issues associated with quality adjustment of prices in a replacement sampling context. Moreover, the notation used in our paper will not be familiar to most practitioners and so Jan has done us all a favor in translating our rather formal matrix algebra results into an easier to interpret framework.

In order to help the reader make the connection between our notation and the notation used by Haan, we will specialize our unweighted models discussed in sections 2 and 3 of our paper to the case where the number of new items that enter the sample in period 1 is equal to the number of items which have disappeared from the sample in period 0 so that the total number of items in the sample in period 0, N(0), is equal to the total number of items or models in period 1, N(1), and we will follow Haan and set n equal to this common number of models. With this replacement sampling simplification of our model, the exponential of LP_{HI} defined by our equation (34), where LP_{HI} is our hedonic imputation estimate of the change in log prices going from period 0 to 1, is indeed equal to Haan’s hedonic imputation index, \( \hat{P}_{HI} \), defined by his equation (4) and as Haan notes, the exponential of our LP_{HI} is also equal to Haan’s full imputation index, \( \hat{P}_{FI} \), defined by his equation (3). Furthermore, using our expressions (32) and (33) and the simplification that N(0) equals N(1), it is easy to show that the exponential of LP_{HI} defined by our equation (34) is also equal to Haan’s double imputation price index, \( \hat{P}_{DI} \), defined by his equation (2). Note that Haan’s double imputation price index uses the actual prices for the matched models and hence using the above equalities, so does our hedonic imputation.
index, $\text{LP}_{\text{HI}}$. Thus the main point to debate in this context is whether to use Haan’s single imputation index $\hat{P}_i$, defined by his equation (1), or the double imputation index which was defined (in logarithms) in our paper by equation (34) and which is equal to Haan’s expressions (2)-(3). For a discussion of the merits of the two methods, the reader is referred to Haan’s commentary.

Haan also briefly discusses our weighted hedonic imputation indexes in his commentary and he provides a much more extensive discussion of the issues associated with weighting in hedonic regressions in Haan (2007) and we recommend this paper to interested readers. The specific point that Haan makes in his commentary about our weighted hedonic imputation index (whose logarithm $\text{LP}_{\text{WHI}}$ is defined by (65) in our paper) is that this index does not satisfy the strong identity test; i.e., if the models are exactly the same in the two periods under consideration and the prices for each model remain unchanged, then the strong identity test asks that the index be equal to unity, no matter what the quantities are. Haan is correct in his assertion; the exponential of our $\text{LP}_{\text{WHI}}$ defined by (65) does not satisfy the strong identity test, whereas his preferred Törnqvist imputation index defined by his equation (13) does satisfy this test. Haan ends his commentary by noting that the issue of weighting in hedonic regressions seems to be unresolved; i.e., is our form of weighting to be preferred over his or not? This issue requires more research but at this point in time, we do find Haan’s suggested weighting scheme rather attractive!

Reference