Volume Title: Price Index Concepts and Measurement

Volume Author/Editor: W. Erwin Diewert, John S. Greenlees and Charles R. Hulten, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-14855-6

Volume URL: http://www.nber.org/books/diew08-1

Conference Date: June 28-29, 2004

Publication Date: December 2009

Chapter Title: Hedonic Imputation versus Time Dummy Hedonic Indexes

Chapter Author: W. Erwin Diewert, Saeed Heravi, Mick Silver

Chapter URL: http://www.nber.org/chapters/c5073

Chapter pages in book: (161 - 196)

4

Hedonic Imputation versus Time Dummy Hedonic Indexes

W. Erwin Diewert, Saeed Heravi, and Mick Silver

4.1 Introduction

The purpose of this chapter 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 chapter 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 4.2 and 4.3. 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 4.4 and 4.5. 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 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.[1] 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 Schultze 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 Schultze 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. (Schultze and 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[2] though to

---

1. See for example, Cole et al. (1986), Silver and Heravi (2003, 2005), Pakes (2003), and Triplett (2004).

2. See Berndt, Griliches, and Rappaport (1995), Berndt and Rappaport (2001), Diewert (2003b), Silver and Heravi (2003), Pakes (2003), de Haan (2003, 2004), and Triplett (2004).

the authors' knowledge no formal analysis has been presented, with a few exceptions,[3] 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 that will be considered in sections 4.5 and 4.7.

Section 4.2 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 4.3, 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.[4] An exact expression for the difference in constant quality log price change between the time dummy and imputation measures is also developed in section 4.3. It is found that in order for these two overall measures to differ, we require the following.

- 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[5]
- Differences in estimated hedonic coefficients for the two separate hedonic regressions

The analysis in sections 4.2 and 4.3 is repeated in the weighted context in sections 4.4 and 4.5. Section 4.6 provides an empirical study for desktop PCs and section 4.7 concludes by discussing the issue of choice between the approaches in light of the theoretical and empirical findings.

Appendix A 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 periods under consideration. The first methodological approach is due to

3. 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 chapter, who developed an expression for the difference based on a framework outlined in Triplett and MacDonald (1977). The points made in Haan's commentary are developed more fully in de Haan (2007).

4. An alternative interpretation of this measure of price change is derived in appendix B.

5. 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).

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 A 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 B, a method for obtaining approximate standard errors for the Laspeyres and Paasche hedonic imputation measures of log price change is derived.

4.2   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, $[z_{n1}^t, z_{n2}^t, \ldots, z_{nK}^t]$ for $t = 0,1$ and $n = 1, 2, \ldots, N(t)$. The time dummy regression model has the following form:

$$(1) \quad \ln p_n^t \equiv y_n^t = \alpha_t + \sum_{k=1}^{K} z_{nk}^t \gamma_k + \varepsilon_n^t; \qquad t = 0, 1 \, ; n = 1, 2, \ldots, N(t)$$

where the ε_n^t are independently distributed normal variables with mean 0 and constant variance and $\alpha_0, \alpha_1, \gamma_1, \ldots, \gamma_K$ are parameters to be estimated. The parameters α_0 and α_1 are measures of the *average level of constant quality prices* of the items in period 0 and 1, respectively, and the $\gamma_1, \ldots, \gamma_K$ are *quality adjustment factors* for the K characteristics; that is, γ_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 we have separate constant terms for each period. The overall measure of *logarithmic price change* going from period 0 to period 1 is $\alpha_1 - \alpha_0$.[6] Let 1_t and 0_t be vectors of ones and zeros of dimension $N(t)$, let $\mathbf{y}^0$ and $\mathbf{y}^1$ be the $N(0)$ and $N(1)$ dimensional vectors of period 0 and 1 logarithms of product prices, let $\boldsymbol{\varepsilon}^0$ and $\boldsymbol{\varepsilon}^1$ be the $N(0)$ and $N(1)$ dimensional vectors of period 0 and 1 stochastic disturbances and let $\mathbf{Z}^0$ and $\mathbf{Z}^1$ be matrices of the product characteristics in periods 0 and 1, respectively. Then the model defined by equation (1) can be written in matrix notation as follows:

$$(2) \qquad \mathbf{y}^0 = 1_0\alpha_0 + 0_0\alpha_1 + \mathbf{Z}^0\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^0;$$

$$(3) \qquad \mathbf{y}^1 = 0_1\alpha_0 + 1_1\alpha_1 + \mathbf{Z}^1\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^1.$$

Let $\alpha_0^*, \alpha_1^*, \gamma_1^*, \ldots, \gamma_K^*$ be the maximum likelihood or least squares estimators for the parameters that appear in equations (2) and (3). Then letting $\mathbf{e}^0$ and $\mathbf{e}^1$ 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:

$$(4) \qquad \mathbf{y}^0 = 1_0\alpha_0^* + 0_0\alpha_1^* + \mathbf{Z}^0\boldsymbol{\gamma}^* + \mathbf{e}^0;$$

$$(5) \qquad \mathbf{y}^1 = 0_1\alpha_0^* + 1_1\alpha_1^* + \mathbf{Z}^1\boldsymbol{\gamma}^* + \mathbf{e}^1.$$

Let $\mathbf{y} \equiv [\mathbf{y}^{0T}, \mathbf{y}^{1T}]^T$ and $\mathbf{e} \equiv [\mathbf{e}^{0T}, \mathbf{e}^{1T}]^T$ and define $\boldsymbol{\phi}^* \equiv [\alpha_0^*, \alpha_1^*, \gamma_1^*, \ldots, \gamma_K^*]^T$. Now rewrite equations (4) and (5) as

$$(6) \qquad \mathbf{y} = X\boldsymbol{\phi}^* + \mathbf{e}.$$

It is well known that the columns of the $\mathbf{X}$ matrix are orthogonal to the vector $\mathbf{e}$ of least squares residuals; that is, we have

$$(7) \qquad \mathbf{X}^T\mathbf{e} = \mathbf{X}^T[\mathbf{y} - \mathbf{X}\boldsymbol{\phi}^*] = 0_{2+K}.$$

The first two equations in (7) are equivalent to the following two equations:[7]

$$(8) \qquad 1_0^T\mathbf{y}^0 = N(0)\alpha_0^* + 1_0^T\mathbf{Z}^0\boldsymbol{\gamma}^*;$$

$$(9) \qquad 1_1^T\mathbf{y}^1 = N(1)\alpha_1^* + 1_1^T\mathbf{Z}^1\boldsymbol{\gamma}^*.$$

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

---

6. 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.

7. Diewert (2003a, 335; 2003b, 39; 2006) 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.

$$(10) \qquad \alpha_0^* = \frac{1_0^T \mathbf{y}^0}{N(0)} - \frac{1_0^T \mathbf{Z}^0 \boldsymbol{\gamma}^*}{N(0)} = \frac{1_0^T (\mathbf{y}^0 - \mathbf{Z}^0 \boldsymbol{\gamma}^*)}{N(0)};$$

$$(11) \qquad \alpha_1^* = \frac{1_1^T \mathbf{y}^1}{N(1)} - \frac{1_1^T \mathbf{Z}^1 \boldsymbol{\gamma}^*}{N(1)} = \frac{1_1^T (\mathbf{y}^1 - \mathbf{Z}^1 \boldsymbol{\gamma}^*)}{N(1)}.$$

Note that $1_t^T \mathbf{y}^t / N(t)$ is the arithmetic average of the log prices in period t for $t = 0,1$. Furthermore, note that $1_0^T \mathbf{Z}^0 / N(0)$ is the arithmetic average of the amounts of each characteristic that are present in the period 0 models and $1_1^T \mathbf{Z}^1 / N(1)$ is the corresponding arithmetic average amount of each characteristic that is present in the period 1 models. Thus, each α_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 $\boldsymbol{\gamma}^*$ 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 α_t^* is equal to the arithmetic average of the quality adjusted log prices, $\mathbf{y}^t - \mathbf{Z}^t \boldsymbol{\gamma}^*$, for the models 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, $\mathrm{LP}_{\mathrm{HD}}$, is the following difference in the log price levels:[8]

$$(12) \qquad \mathrm{LP}_{\mathrm{HD}} \equiv \alpha_1^* - \alpha_0^*.$$

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

$$(13) \qquad \mathbf{X} = [\mathbf{V}, \mathbf{Z}],$$

where $\mathbf{Z}^T \equiv [\mathbf{Z}^{0T}, \mathbf{Z}^{1T}]$ and $\mathbf{V}$ is an $N(0) + N(1)$ by 2 matrix that 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 equation (6) in two stages. In the first stage, we condition on γ and minimize with respect to the components of $\alpha \equiv [\alpha_0, \alpha_1]^T$. The resulting conditional least squares estimator for α is:

$$(14) \qquad \alpha(\gamma) \equiv (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T (y - Z\gamma).$$

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

---

8. 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 $N(0)$ equals $N(1)$ (equals N, say) and $\mathbf{Z}^0$ equals $\mathbf{Z}^1$, so that the two characteristics matrices are identical and thus $\alpha_1^* - \alpha_0^* = 1_1^T[\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*]/N(1) - 1_0^T[\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*]/N(0) = 1_1^T\mathbf{y}^1/N - 1^T\mathbf{y}^0/N$, 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.

(15) $f(\boldsymbol{\gamma}) \equiv [\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma} - \mathbf{V}\alpha(\boldsymbol{\gamma})]^T[\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma} - \mathbf{V}\alpha(\boldsymbol{\gamma})]$

$\qquad = [\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma} - \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})]^T$

$\qquad\quad [\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma} - \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})]$ $\qquad$ using (14)

$\qquad = (\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{Z}\boldsymbol{\gamma})^T(\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{Z}\boldsymbol{\gamma})$

$\qquad = (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})^T\mathbf{M}^T\mathbf{M}(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})$

$\qquad = (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})^T\mathbf{M}(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})$ $\qquad$ since $\mathbf{M} = \mathbf{M}^T$ and $\mathbf{M}^2 = \mathbf{M}$

where the projection matrix $\mathbf{M}$ is defined as follows:

(16) $\qquad\qquad\qquad \mathbf{M} \equiv I - \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T.$

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

(17) $\qquad\qquad\qquad \mathbf{y}^* \equiv \mathbf{M}y; \mathbf{Z}^* \equiv \mathbf{M}Z.$

Using definitions (17), it can be seen that

(18) $\qquad\qquad\qquad f(\boldsymbol{\gamma}) = (\mathbf{y}^* - \mathbf{Z}^*\boldsymbol{\gamma})^T(\mathbf{y}^* - \mathbf{Z}^*\boldsymbol{\gamma}).$

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

(19) $\qquad\qquad\qquad \boldsymbol{\gamma}^* \equiv (\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*T}\mathbf{y}^*.$

Once $\boldsymbol{\gamma}^*$ has been determined by equation (19), then we can use equations (14) or (8) and (9) to determine the least squares estimators for α_0^* and α_1^*.

Using the definition of $\mathbf{V}$, it can be shown that the projection matrix $\mathbf{M}$ defined by (16) is block diagonal, with the two main diagonal blocks $\mathbf{M}^0$ and $\mathbf{M}^1$ defined as follows:

(20) $\qquad\qquad \mathbf{M}^0 \equiv \mathbf{I}_0 - \dfrac{1_0 1_0^T}{N(0)}; \quad \mathbf{M}^1 \equiv \mathbf{I}_1 - \dfrac{1_1 1_1^T}{N(1)},$

where $\mathbf{I}_0$ and $\mathbf{I}_1$ are identity matrices of dimension $N(0)$ and $N(1)$, respectively. Using (20), we can determine more precisely what the vector $\mathbf{y}^*$ equal to $\mathbf{M}y$ and the matrix $\mathbf{Z}^*$ equal to $\mathbf{M}Z$ look like. Let $\mathbf{y}^{*T} = [\mathbf{y}^{0*T}, \mathbf{y}^{1*T}]$ and $\mathbf{Z}^{*T} = [\mathbf{Z}^{0*T}, \mathbf{Z}^{1*T}]$. Then using (20), we have:

(21) $\qquad\qquad \mathbf{y}^{t*} = \mathbf{y}^t - \dfrac{1_t 1_t^T \mathbf{y}^t}{N(t)}; \qquad t = 0, 1;$

(22) $\qquad\qquad \mathbf{Z}^{t*} = \mathbf{Z}^t - \dfrac{1_t 1_t^T \mathbf{Z}^t}{N(t)}; \qquad t = 0, 1.$

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

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

$$(23) \qquad \boldsymbol{\gamma}^* \equiv (\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}(\mathbf{Z}^{0*T}\mathbf{y}^{0*} + \mathbf{Z}^{1*T}\mathbf{y}^{1*}).$$

We now turn our attention to hedonic imputation indexes.

4.3   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, $\boldsymbol{\gamma}^0$ in period 0 and $\boldsymbol{\gamma}^1$ in period 1. Thus using the same notation as in section 4.2, our models now are:

$$(24) \qquad\qquad \mathbf{y}^0 = 1_0 \beta_0 + \mathbf{Z}^0 \boldsymbol{\gamma}^0 + \boldsymbol{\eta}^0;$$

$$(25) \qquad\qquad \mathbf{y}^1 = 1_1 \beta_1 + \mathbf{Z}^1 \boldsymbol{\gamma}^1 + \boldsymbol{\eta}^1,$$

where $\boldsymbol{\eta}^0$ and $\boldsymbol{\eta}^1$ are independently distributed normal random variables with means zero and constant variance within each period. Let $\beta_0^*, \gamma_1^{0*}, \ldots,$ γ_K^{0*} be the maximum likelihood or least squares estimators for the parameters that appear in equation (24) and let $\beta_1^*, \gamma_1^{1*}, \ldots, \gamma_K^{1*}$ be the maximum likelihood or least squares estimators for the parameters that appear in equation (25). Then letting $\mathbf{u}^0$ and $\mathbf{u}^1$ 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:

$$(26) \qquad\qquad \mathbf{y}^0 = 1_0 \beta_0^* + \mathbf{Z}^0 \boldsymbol{\gamma}^{0*} + \mathbf{u}^0;$$

$$(27) \qquad\qquad \mathbf{y}^1 = 1_1 \beta_1^* + \mathbf{Z}^1 \boldsymbol{\gamma}^{1*} + \mathbf{u}^1.$$

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

$$(28) \qquad\qquad 1_0^T \mathbf{y}^0 = N(0)\beta_0^* + 1_0^T \mathbf{Z}^0 \boldsymbol{\gamma}^{0*};$$

$$(29) \qquad\qquad 1_1^T \mathbf{y}^1 = N(1)\beta_1^* + 1_1^T \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}.$$

Equations (28) and (29) lead to the following counterparts to equations (8) and (9):

$$(30) \qquad \beta_0^* = \frac{1_0^T \mathbf{y}^0}{N(0)} - \frac{1_0^T \mathbf{Z}^0 \boldsymbol{\gamma}^{0*}}{N(0)} = \frac{1_0^T (\mathbf{y}^0 - \mathbf{Z}^0 \boldsymbol{\gamma}^{0*})}{N(0)};$$

$$(31) \qquad \beta_1^* = \frac{1_1^T \mathbf{y}^1}{N(1)} - \frac{1_1^T \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}}{N(1)} = \frac{1_1^T (\mathbf{y}^1 - \mathbf{Z}^1 \boldsymbol{\gamma}^{1*})}{N(1)}.$$

Recall that the hedonic time dummy estimate of the change in log prices going from period 0 to 1, $\mathrm{LP_{HD}}$, 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 β_1^* and β_0^* as a measure of constant quality log price change between periods 0 and 1, because the quality adjustment parameters, $\boldsymbol{\gamma}^{0*}$ and $\boldsymbol{\gamma}^{1*}$, are *different* between the two periods. However, we can use the period 0 parameters, $\boldsymbol{\gamma}^{0*}$, 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 δ_1^*:

$$(32) \qquad \delta_1^* \equiv \frac{1_1^T \mathbf{y}^1}{N(1)} - \frac{1_1^T \mathbf{Z}^1 \boldsymbol{\gamma}^{0*}}{N(1)} = \frac{1_1^T (\mathbf{y}^1 - \mathbf{Z}^1 \boldsymbol{\gamma}^{0*})}{N(1)}.$$

Note that the previous estimate of a period 1 log price level is analogous to β_1^*, defined by equation (31) except that the period 0 hedonic quality adjustment factors, $\boldsymbol{\gamma}^{0*}$, are used in equation (32) whereas the period 1 hedonic quality adjustment factors, $\boldsymbol{\gamma}^{1*}$, were used in equation (31). Since the period 0 and 1 estimated price levels, β_0^* and δ_1^*, use the same quality adjustment factors $\boldsymbol{\gamma}^{0*}$ in order to form constant quality log prices in each period, we can take the difference δ_1^* less β_0^* as a measure of quality-adjusted log price change between periods 0 and 1.[9] We call this hedonic imputation measure of log price change a *Laspeyres type measure* of price change and denote it by $\phi_L^* \equiv \delta_1^* - \beta_0^*$. This measure of overall log price change depends asymmetrically on the characteristics price vector $\boldsymbol{\gamma}^{0*}$ 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 $\boldsymbol{\gamma}^{1*}$. Thus, use the period 1 characteristics price vector, $\boldsymbol{\gamma}^{1*}$, 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 δ_0^*:

$$(33) \qquad \delta_0^* \equiv \frac{1_0^T \mathbf{y}^0}{N(0)} - \frac{1_0^T \mathbf{Z}^0 \boldsymbol{\gamma}^{1*}}{N(0)} = \frac{1_0^T (\mathbf{y}^0 - \mathbf{Z}^0 \boldsymbol{\gamma}^{1*})}{N(0)}.$$

Note that the previous estimate of a period 0 log price level is analogous to β_0^* defined by equation (30) except that the period 1 hedonic quality adjustment factors, $\boldsymbol{\gamma}^{1*}$, are used in equation (33), whereas the period 0 hedonic quality adjustment factors, $\boldsymbol{\gamma}^{0*}$, were used in equation (30). Since the period 0 and 1 estimated price levels, δ_0^* and β_1^*, use the same quality adjustment factors $\boldsymbol{\gamma}^{1*}$ in order to form constant quality log prices in each

---

9. 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).

period, we can take the difference β_1^* less δ_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 $\phi_P^* \equiv \beta_1^* - \delta_0^*$.[10]

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[11] 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:[12]

$$
(34) \quad
\begin{aligned}
LP_{HI} &\equiv \frac{1}{2}\phi_L^* + \frac{1}{2}\phi_P^* \\[2mm]
&= \frac{1}{2}(\delta_1^* - \beta_0^*) + \frac{1}{2}(\beta_1^* - \delta_0^*) \\[2mm]
&= \frac{1}{2}\left[\frac{1_1^T(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{0*})}{N(1)} - \frac{1_0^T(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^{0*})}{N(0)}\right. \\[2mm]
&\qquad \left. + \frac{1_1^T(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{1*})}{N(1)} - \frac{1_0^T(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^{1*})}{N(0)}\right] \quad \text{using } (30)-(33) \\[2mm]
&= 1_1^T\left\{\frac{\mathbf{y}^1 - \mathbf{Z}^1[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]}{N(1)}\right\} \\[2mm]
&\quad - 1_0^T\left\{\frac{\mathbf{y}^0 - \mathbf{Z}^0[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]}{N(0)}\right\}.
\end{aligned}
$$

Recall that in the hedonic time dummy method for quality adjusting log prices $\mathbf{y}^t$ for each period t, we used the characteristics quality adjustments defined by $\mathbf{Z}^t\boldsymbol{\gamma}^*$, where $\boldsymbol{\gamma}^*$ was a constant across periods vector of quality adjustment factors. Looking at the right-hand side of (34), it can be seen that the hedonic imputation method for quality adjusting log prices in each period is similar but now the period t vector of quality adjustments is $\mathbf{Z}^t[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]$ instead of $\mathbf{Z}^t\boldsymbol{\gamma}^*$. Thus for the hedonic imputation method of quality adjusting prices, the average of the two separate hedonic

---

10. In appendix B 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.

11. 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 $\exp[\beta_1^* - \delta_0^*]$.

12. In appendix A we show that the measure of quality-adjusted change in log prices defined by equation (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, 111–12) and further developed by Feenstra (1995) and Diewert (2003a, 341–44). Empirical studies using this approach include Berndt, Griliches, and Rappaport (1995), Ioannidis and Silver (1999), Berndt and Rappaport (2001, 270), Koskimäki and Vartia (2001) and Silver and Heravi (2003, 2007b).

regression estimated quality adjustment factors, $(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}$, replaces the single regression estimated quality adjustment factor vector $\boldsymbol{\gamma}^*$ that was used in the hedonic time dummy method for quality adjustment.

Note also that if the models are the same in the two periods being considered, then $N(0)$ equals $N(1)$ (equals N, say) and $\mathbf{Z}^0$ equals $\mathbf{Z}^1$, so that the two characteristics matrices are identical, then LP_{HI} defined by (34) collapses to $1_1^T\mathbf{y}^1/N - 1^T\mathbf{y}^0/N$, 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 hedonic regressions; the usual matched-model methodology can be used.

Using (12) and (34), we can form the following expression for the difference in the overall log price change using the two methods for quality adjustment:

$$(35)\quad LP_{HD} - LP_{HI} = -\frac{1_1^T\mathbf{Z}^1\boldsymbol{\gamma}^*}{N(1)} + \frac{1_0^T\mathbf{Z}^0\boldsymbol{\gamma}^*}{N(0)} + \frac{1_1^T\mathbf{Z}^1[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]}{N(1)}$$

$$-\frac{1_0^T\mathbf{Z}^0[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]}{N(0)}$$

$$= \frac{1_1^T\mathbf{Z}^1[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*]}{N(1)}$$

$$-\frac{1_0^T\mathbf{Z}^0[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*]}{N(0)}$$

$$= \left[\frac{1_1^T\mathbf{Z}^1}{N(1)} - \frac{1_0^T\mathbf{Z}^0}{N(0)}\right]\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*\right).$$

Thus, the hedonic time dummy and hedonic imputation measures of log price change will be identical if *either* of the following two conditions are satisfied:

$$(36)\qquad\qquad \frac{1_1^T\mathbf{Z}^1}{N(1)} = \frac{1_0^T\mathbf{Z}^0}{N(0)};$$

$$(37)\qquad\qquad \boldsymbol{\gamma}^* = \frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}.$$

Condition (36) says that the average amount of each characteristic for the models present in period 1 equals the corresponding average amount of each characteristic for the models present in period 0. Condition (37) says that the time dummy vector of quality adjustment factors, $\boldsymbol{\gamma}^*$, is equal to the arithmetic average of the two separate hedonic regression estimates for the quality adjustment factors, $(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}$.

Condition (36) is somewhat unanticipated. It tells us that if the average amount of characteristics present in the models in each period does not change much, then the hedonic time dummy and hedonic imputation esti-

mates of quality-adjusted price change will be much the same, even if characteristic valuations change over the two periods.

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, γ^*. 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 γ^{0*} and γ^{1*} that appear in (26) and (27):

$$(38)\qquad \gamma^{0*} = (\mathbf{Z}^{0*T}\mathbf{Z}^{0*})^{-1}\mathbf{Z}^{0*T}\mathbf{y}^{0*}; \; \gamma^{1*} = (\mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}\mathbf{Z}^{1*T}\mathbf{y}^{1*},$$

where the $\mathbf{y}^{t*}$ and $\mathbf{Z}^{t*}$ are the demeaned $\mathbf{y}^t$ and $\mathbf{Z}^t$ as in the previous section.[13] Now premultiply both sides of (23) by the matrix $\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*}$ and we obtain the following equation:[14]

$$
\begin{aligned}
(39)\quad (\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})\gamma^* &= \mathbf{Z}^{0*T}\mathbf{y}^{0*} + \mathbf{Z}^{1*T}\mathbf{y}^{1*} \\
&= \mathbf{Z}^{0*T}\mathbf{Z}^{0*}(\mathbf{Z}^{0*T}\mathbf{Z}^{0*})^{-1}\mathbf{Z}^{0*T}\mathbf{y}^{0*} \\
&\quad + \mathbf{Z}^{1*T}\mathbf{Z}^{1*}(\mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}\mathbf{Z}^{1*T}\mathbf{y}^{1*} \\
&= \mathbf{Z}^{0*T}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*}\gamma^{1*} \quad \text{using (38).}
\end{aligned}
$$

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

$$
\begin{aligned}
(40)\quad 2(\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})&\left[\frac{1}{2}\gamma^{0*} + \frac{1}{2}\gamma^{1*} - \gamma^*\right] \\
&= (\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})(\gamma^{0*} + \gamma^{1*}) \\
&\quad - 2(\mathbf{Z}^{0*T}\mathbf{Z}^{0*}\gamma^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*}\gamma^{1*}) \quad \text{using (39).} \\
&= (\mathbf{Z}^{0*T}\mathbf{Z}^{0*})(\gamma^{1*} - \gamma^{0*}) - (\mathbf{Z}^{1*T}\mathbf{Z}^{1*})(\gamma^{1*} - \gamma^{0*}) \\
&= -(\mathbf{Z}^{1*T}\mathbf{Z}^{1*} - \mathbf{Z}^{0*T}\mathbf{Z}^{0*})(\gamma^{1*} - \gamma^{0*})
\end{aligned}
$$

Now premultiply both sides of (40) by $(1/2)(\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}$ and substitute the resulting expression for $(1/2)\gamma^{0*} + (1/2)\gamma^{1*} - \gamma^*$ 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:

$$
\begin{aligned}
(41)\quad \mathrm{LP_{HD}} - \mathrm{LP_{HI}} = &-\frac{1}{2}\left[\frac{\mathbf{1}_1^T\mathbf{Z}^1}{N(1)} - \frac{\mathbf{1}_0^T\mathbf{Z}^0}{N(0)}\right] \\
&(\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}(\mathbf{Z}^{1*T}\mathbf{Z}^{1*} - \mathbf{Z}^{0*T}\mathbf{Z}^{0*}) \\
&(\gamma^{1*} - \gamma^{0*}).
\end{aligned}
$$

13. Note that $\mathbf{Z}^{t*T}\mathbf{y}^{t*}/N(t)$ 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 $\mathbf{Z}^{t*T}\mathbf{Z}^{t*}/N(t)$ can be interpreted as a sample variance covariance matrix for the model characteristics in period t. In the main text, we refer to $\mathbf{Z}^{t*T}\mathbf{Z}^{t*}$ as a characteristics "total" variance covariance matrix.

14. In the single characteristic case, equation (40) tells us that γ_1^* is a weighted average of γ_1^{0*} and γ_1^{1*}.

Using equation (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:

- $1_1^T\mathbf{Z}^1/N(1)$ equals $1_0^T\mathbf{Z}^0/N(0)$ so that the average amount of each characteristic across models in each period stays the same
- $\mathbf{Z}^{1*T}\mathbf{Z}^{1*}$ equals $\mathbf{Z}^{0*T}\mathbf{Z}^{0*}$ so that the model characteristics total variance covariance matrix is the same across periods[15]
- $\boldsymbol{\gamma}^{1*}$ equals $\boldsymbol{\gamma}^{0*}$ so that separate (unweighted) hedonic regressions in each period give rise to the same characteristics quality adjustment factors.[16]

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

4.4   Weighted Time Dummy Hedonic Regressions

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_n^t, 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_n^t, are also observable. Model n in period t again has the vector of K characteristics associated with it, $[z_{n1}^t, z_{n2}^t, \ldots, z_{nK}^t]$ for $t = 0, 1$ and $n = 1, 2, \ldots, N(t)$. The *expenditure share* of model n in period t is

(42)
$$\mathbf{s}_n^t \equiv \frac{p_n^t q_n^t}{\sum_{i=1}^{N(t)} p_i^t q_i^t}; \qquad t = 0, 1; n = 1, 2, \ldots, N(t).$$

Let $\mathbf{s}^t \equiv [s_1^t, \ldots, s_{N(t)}^t]^T$ denote the period t expenditure share vector for $t = 0, 1$ and let $\mathbf{S}^t$ denote the *diagonal period t share matrix* that has the elements of the period t expenditure share vector $\mathbf{s}^t$ running down the main diagonal for $t = 0,1$. The matrix $(\mathbf{S}^t)^{1/2}$ is the *square root matrix* of $\mathbf{S}^t$; that is, the positive square roots of the elements of the period t expenditure vector $\mathbf{s}^t$ are the diagonal elements of this diagonal matrix for $t = 0, 1$. As in the previous sections, $\mathbf{y}^t$ is the vector of period t log price changes and $\mathbf{Z}^t$ is the period t model characteristics matrix for $t = 0,1$.

With the previous 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:[17]

---

15. This condition for equality is also somewhat unanticipated.

16. Using equation (38), we can obtain a more "fundamental" condition in terms of variance covariance matrices for the equality of $\boldsymbol{\gamma}^{0*}$ to $\boldsymbol{\gamma}^{1*}$; namely the equality of $(\mathbf{Z}^{0*T}\mathbf{Z}^{0*})^{-1}\mathbf{Z}^{0*T}\mathbf{y}^{0*}$ to $(\mathbf{Z}^{1*T}\mathbf{Z}^{1*})^{-1}\mathbf{Z}^{1*T}\mathbf{y}^{1*}$.

17. Diewert (2003b, 26) explained the logic behind the weighting scheme in the regression model defined by equations (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 equations (43) and (44) leads to measures of price change that are comparable (in the case where

(43) $\qquad (\mathbf{S}^0)^{1/2}\mathbf{y}^0 = (\mathbf{S}^0)^{1/2}(1_0\alpha_0 + 0_0\alpha_1 + \mathbf{Z}^0\boldsymbol{\gamma}) + \boldsymbol{\varepsilon}^0;$

(44) $\qquad (\mathbf{S}^1)^{1/2}\mathbf{y}^1 = (\mathbf{S}^1)^{1/2}(0_1\alpha_0 + 1_1\alpha_1 + \mathbf{Z}^1\boldsymbol{\gamma}) + \boldsymbol{\varepsilon}^1,$

where the $\boldsymbol{\varepsilon}^t$ vectors have elements ε_n^t that are independently distributed normal variables with zero means and constant variances.

Let $\alpha_0^*, \alpha_1^*, \gamma_1^*, \ldots, \gamma_K^*$ be the maximum likelihood or least squares estimators for the parameters that appear in (43) and (44). Then letting $\mathbf{e}^0$ and $\mathbf{e}^1$ 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) $\qquad (\mathbf{S}^0)^{1/2}\mathbf{y}^0 = (\mathbf{S}^0)^{1/2}(1_0\alpha_0^* + 0_0\alpha_1^* + \mathbf{Z}^0\boldsymbol{\gamma}^*) + \mathbf{e}^0;$

(46) $\qquad (\mathbf{S}^1)^{1/2}\mathbf{y}^1 = (\mathbf{S}^1)^{1/2}(0_1\alpha_0^* + 1_1\alpha_1^* + \mathbf{Z}^1\boldsymbol{\gamma}^*) + \mathbf{e}^1.$

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

(47) $\qquad \alpha_0^* = \mathbf{s}^{0T}\mathbf{y}^0 - \mathbf{s}^{0T}\mathbf{Z}^0\boldsymbol{\gamma}^* = \mathbf{s}^{0T}(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*);$

(48) $\qquad \alpha_1^* = \mathbf{s}^{1T}\mathbf{y}^1 - \mathbf{s}^{1T}\mathbf{Z}^1\boldsymbol{\gamma}^* = \mathbf{s}^{1T}(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*).$

Note that $\mathbf{s}^{tT}\mathbf{y}^t = \Sigma_{n=1}^{N(t)} s_n^t \ln p_n^t$ is the period t expenditure share weighted average of the log prices in period t for $t = 0,1$. Furthermore, note that $\mathbf{s}^{tT}\mathbf{Z}^t$ is a 1 by K vector whose kth element is equal to the period t expenditure share weighted average $\Sigma_{n=1}^{N(t)} s_n^t z_{nk}^t$ of the amounts of characteristic k that are present in the period t models for $t = 0, 1$ and $k = 1, \ldots, K$. Thus, each α_t^* 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 $\boldsymbol{\gamma}^*$ 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 α_t^* is equal to the period t expenditure share weighted average of the quality-adjusted log prices, $\mathbf{y}^t - \mathbf{Z}^t\boldsymbol{\gamma}^*$, for the models present in that period. Now use equations (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: LP_{WHD}, is the following difference in the log price levels:

(49) $\qquad LP_{WHD} \equiv \alpha_1^* - \alpha_0^* = \mathbf{s}^{1T}(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*) - \mathbf{s}^{0T}(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*).$

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, $\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*$, less a period 0 expenditure share weighted average of the quality-adjusted log prices, $\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*$.

---

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_n^t is equal to a common value.

For later reference, we can adapt the methodology presented at the end of section 4.2 in order to work out an explicit expression for the estimated characteristic prices, $\boldsymbol{\gamma}^*$. Recall the definitions of the diagonal matrices $\mathbf{S}^0$ and $\mathbf{S}^1$. Define $\mathbf{S}$ as a block diagonal matrix that has the blocks $\mathbf{S}^0$ and $\mathbf{S}^1$ on the main diagonals and let $\mathbf{S}^{1/2}$ be the corresponding square root matrix; that is, this matrix has the elements of the period 0 expenditure share $\mathbf{s}^0$ and the period 1 expenditure share vector $\mathbf{s}^1$ running down the main diagonal. Adapting the analysis in section 4.1, we need to replace the $\mathbf{y}$ vector in that section by $\mathbf{S}^{1/2}\mathbf{y}$ and the $\mathbf{Z}$ matrix in that section by $\mathbf{S}^{1/2}\mathbf{Z}$, the $\mathbf{V}$ matrix by $\mathbf{S}^{1/2}\mathbf{V}$, and define the counterpart $\mathbf{M}°$ to the projection matrix $\mathbf{M}$ defined by equation (16) as follows:

$$(50) \qquad \mathbf{M}° \equiv \mathbf{I} - \mathbf{S}^{1/2}\mathbf{V}(\mathbf{V}^T\mathbf{SV})^{-1}\mathbf{V}^T\mathbf{S}^{1/2}$$
$$= \mathbf{I} - \mathbf{S}^{1/2}\mathbf{VV}^T\mathbf{S}^{1/2},$$

where the second equality in equation (50) follows from the fact that $\mathbf{V}^T\mathbf{SV}$ equals $\mathbf{I}_2$, a two-by-two identity matrix. Now define $\mathbf{y}°$ and $\mathbf{Z}°$ in terms of $\mathbf{M}°$ and the original $\mathbf{y}$ vector and $\mathbf{Z}$ matrix as follows:

$$(51) \qquad \mathbf{y}° \equiv \mathbf{M}°\mathbf{S}^{1/2}\mathbf{y}$$
$$= (\mathbf{I} - \mathbf{S}^{1/2}VV^T\mathbf{S}^{1/2})\mathbf{S}^{1/2}\mathbf{y} \qquad \text{using (50);}$$
$$= \mathbf{S}^{1/2}(\mathbf{I} - \mathbf{VV}^T\mathbf{S})\mathbf{y};$$

$$(52) \qquad \mathbf{Z}° \equiv \mathbf{M}°\mathbf{S}^{1/2}\mathbf{Z}$$
$$= (\mathbf{I} - \mathbf{S}^{1/2}\mathbf{VV}^T\mathbf{S}^{1/2})\mathbf{S}^{1/2}\mathbf{Z} \qquad \text{using (50).}$$
$$= \mathbf{s}^{1/2}(\mathbf{I} - \mathbf{VV}^T\mathbf{S})\mathbf{Z}$$

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

$$(53) \qquad \boldsymbol{\gamma}^* \equiv (\mathbf{Z}°^T\mathbf{Z}°)^{-1}\mathbf{Z}°^T\mathbf{y}°.$$

Once $\boldsymbol{\gamma}^*$ has been determined by equation (53), then we can use equations (47) and (48) to determine the weighted least squares estimators for α_0^* and α_1^*.

It is possible to express the $\boldsymbol{\gamma}^*$ defined by (53) in a more transparent way using our definitions of the matrices $\mathbf{Z}$, $\mathbf{V}$, and $\mathbf{S}$. Recall that in section 4.2, the demeaned log price change vectors $\mathbf{y}^{t*}$ defined by (20) and the demeaned characteristics matrices $\mathbf{Z}^{t*}$ 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:

$$(54) \qquad \mathbf{y}^{t*} \equiv \mathbf{y}^t - 1_t\mathbf{s}^{tT}\mathbf{y}^t; \qquad t = 0, 1;$$

$$(55) \qquad \mathbf{Z}^{t*} \equiv \mathbf{Z}^t - 1_t\mathbf{s}^{tT}\mathbf{Z}^t; \qquad t = 0, 1.$$

Let the nth component of the vector $\mathbf{y}^{t*}$ be $\mathbf{y}_n^{t*}$ and let the nth row of the $N(t)$ by K dimensional matrix $\mathbf{Z}^{t*}$ be z_n^{t*} for $t = 0, 1$ and $n = 1, 2, \ldots, N(t)$. Then it can be shown that the vector of characteristics prices $\boldsymbol{\gamma}^*$ defined by (53) can be written in terms of the components of the demeaned log price

change vectors $\mathbf{y}^{t*}$, the components of the expenditure share vectors $\mathbf{s}^t$, and the rows of the demeaned characteristics matrices $\mathbf{Z}^{t*}$ as follows:

$$(56) \qquad \boldsymbol{\gamma}^* = \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} \right]^{-1}$$

$$\left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} y_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} y_n^{1*} \right].$$

Note that $\Sigma_{n=1}^{N(t)} s_n^t z_n^{t*T} z_n^{t*}$ can be interpreted as a period t expenditure share weighted sample variance covariance matrix[18] for the model characteristics present in period t and $\Sigma_{n=1}^{N(t)} s_n^t z_n^{t*T} y_n^{t*}$ 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 4.3 to the weighted case.

4.5   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 equations (57) and (58):

$$(57) \qquad (\mathbf{S}^0)^{1/2} \mathbf{y}^0 = (\mathbf{S}^0)^{1/2} (\mathbf{1}_0 \beta_0 + \mathbf{Z}^0 \boldsymbol{\gamma}^0) + \boldsymbol{\eta}^0;$$

$$(58) \qquad (\mathbf{S}^1)^{1/2} \mathbf{y}^1 = (\mathbf{S}^1)^{1/2} (\mathbf{1}_1 \beta_1 + \mathbf{Z}^1 \boldsymbol{\gamma}^1) + \boldsymbol{\eta}^1,$$

where $\boldsymbol{\eta}^0$ and $\boldsymbol{\eta}^1$ are independently distributed normal random variables with means zero and constant variance within each period. Let $\beta_0^*, \gamma_1^{0*}, \ldots, \gamma_K^{0*}$ be the maximum likelihood or least squares estimators for the parameters that appear in equation (57) and let $\beta_1^*, \gamma_1^{1*}, \ldots, \gamma_K^{1*}$ be the maximum likelihood or least squares estimators for the parameters that appear in equation (58). Then letting $\mathbf{u}^0$ and $\mathbf{u}^1$ 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:

$$(59) \qquad (\mathbf{S}^0)^{1/2} \mathbf{y}^0 = (\mathbf{S}^0)^{1/2} (\mathbf{1}_0 \beta_0^* + \mathbf{Z}^0 \boldsymbol{\gamma}^{0*}) + \mathbf{u}^0;$$

$$(60) \qquad (\mathbf{S}^1)^{1/2} \mathbf{y}^1 = (\mathbf{S}^1)^{1/2} (\mathbf{1}_1 \beta_1^* + \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}) + \mathbf{u}^1.$$

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

$$(61) \qquad \beta_0^* = \mathbf{s}^{0T} \mathbf{y}^0 - \mathbf{s}^{0T} \mathbf{Z}^0 \boldsymbol{\gamma}^{0*} = \mathbf{s}^{0T} (\mathbf{y}^0 - \mathbf{Z}^0 \boldsymbol{\gamma}^{0*});$$

$$(62) \qquad \beta_1^* = \mathbf{s}^{1T} \mathbf{y}^1 - \mathbf{s}^{1T} \mathbf{Z}^1 \boldsymbol{\gamma}^{1*} = \mathbf{s}^{1T} (\mathbf{y}^0 - \mathbf{Z}^1 \boldsymbol{\gamma}^{1*}).$$

18. Note that for the present weighted model, $\Sigma_{n=1}^{N(t)} s_n^t z_n^{t*T} z_n^{t*}$ is a true period t expenditure share weighted sample variance covariance matrix, whereas for the earlier unweighted model, the counterpart to this sample variance covariance matrix was $\mathbf{Z}^{t*T} \mathbf{Z}^{t*}$, which was the period t total variance covariance matrix, equal to $N(t)$ times the sample variance covariance matrix for the characteristics present in period t models.

As in section 4.3, we cannot simply take the difference between β_1^* and β_0^* as a measure of constant quality log price change between periods 0 and 1, because the quality adjustment parameters, γ^{0*} and γ^{1*}, are again *different* between the two periods. As in section 4.3, we can use the period 0 parameters, γ^{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 δ_1^*:

$$(63) \qquad \delta_1^* \equiv \mathbf{s}^{1T}\mathbf{y}^1 - \mathbf{s}^{1T}\mathbf{Z}^1\gamma^{0*} = \mathbf{s}^{1T}(\mathbf{y}^1 - \mathbf{Z}^1\gamma^{0*}).$$

Note that the previous estimate of a period 1 log price level is analogous to β_1^* defined by equation (62) except that the period 0 hedonic quality adjustment factors, γ^{0*}, are used in equation (63), whereas the period 1 hedonic quality adjustment factors, γ^{1*}, were used in equation (62). Since the period 0 and 1 estimated price levels, β_0^* and δ_1^*, use the same quality adjustment factors γ^{0*} in order to form constant quality log prices in each period, we can again take the difference δ_1^* less β_0^* as a measure of quality-adjusted log price change between periods 0 and 1.[19] This measure of overall log price change depends asymmetrically on the characteristics price vector γ^{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 γ^{1*}. Thus, use the period 1 characteristics price vector, γ^{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 δ_0^*:

$$(64) \qquad \delta_0^* \equiv \mathbf{s}^{0T}\mathbf{y}^0 - \mathbf{s}^{0T}\mathbf{Z}^0\gamma^{1*} = \mathbf{s}^{0T}(\mathbf{y}^0 - \mathbf{Z}^0\gamma^{1*}).$$

Note that the previous estimate of a period 0 log price level is analogous to β_0^* defined by equation (61) except that the period 1 hedonic quality adjustment factors, γ^{1*}, are used in (64) whereas the period 0 hedonic quality adjustment factors, γ^{0*}, were used in equation (61). Since the period 0 and 1 estimated price levels, δ_0^* and β_1^*, use the same quality adjustment factors γ^{1*} in order to form constant quality log prices in each period, we can take the difference β_1^* less δ_0^* as a second measure of log price change between periods 0 and 1.[20]

Again following Diewert (2003b, 20) and de Haan (2003, 14; 2004), it seems preferable to take a symmetric average of the two previous estimates of log price change over the two periods. We again choose the arithmetic mean

19. 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.

20. 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.

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:[21]

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

$$= \frac{1}{2}[\mathbf{s}^{1T}(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{0*}) - \mathbf{s}^{0T}(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^{0*})$$
$$+ \mathbf{s}^{1T}(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{1*}) - \mathbf{s}^{0T}(\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^{1*})]$$

$$= \mathbf{s}^{1T}\left[\mathbf{y}^1 - \mathbf{Z}^1\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right]$$
$$- \mathbf{s}^{0T}\left[\mathbf{y}^0 - \mathbf{Z}^0\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right].$$

Recall that in the weighted hedonic time dummy method for quality adjusting log prices $\mathbf{y}^t$ for each period t, we used the characteristics quality adjustments defined by $\mathbf{Z}^t\boldsymbol{\gamma}^*$, where $\boldsymbol{\gamma}^*$ was a constant across periods vector of quality adjustment factors. Looking at the right-hand side of equation (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 $\mathbf{Z}^t[(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*}]$ instead of $\mathbf{Z}^t\boldsymbol{\gamma}^*$.

Using equations (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:

$$(66) \qquad LP_{WHD} - LP_{WHI} = \mathbf{s}^{1T}[\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^*] - \mathbf{s}^{0T}[\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^*]$$
$$- \left\{\mathbf{s}^{1T}\left[\mathbf{y}^1 - \mathbf{Z}^1\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right]\right.$$
$$\left. - \mathbf{s}^{0T}\left[\mathbf{y}^0 - \mathbf{Z}^0\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}\right)\right]\right\}$$
$$= (\mathbf{s}^{1T}\mathbf{Z}^1 - \mathbf{s}^{0T}\mathbf{Z}^0)\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*\right).$$

Thus, the hedonic time dummy and hedonic imputation measures of log price change will be identical if *either* of the following two conditions are satisfied:

$$(67) \qquad \mathbf{s}^{1T}\mathbf{Z}^1 = \mathbf{s}^{0T}\mathbf{Z}^0;$$

$$(68) \qquad \boldsymbol{\gamma}^* = \frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*}.$$

---

21. 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.

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, γ^*, is equal to the arithmetic average of the two separate weighted hedonic regression estimates for the quality adjustment factors, $(1/2)\gamma^{0*} + (1/2)\gamma^{1*}$.

As in section 4.3, condition (68) can be strengthened. Recall equation (56) in the previous section, which provided a formula for the hedonic time dummy vector of quality adjustment factors, γ^*. 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 γ^{0*} and γ^{1*} that appear in equations (59) and (60):

$$(69) \qquad \gamma^{0*} = \left(\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} \right)^{-1} \left(\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} y_n^{0*} \right);$$

$$(70) \qquad \gamma^{1*} = \left(\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} \right)^{-1} \left(\sum_{n=1}^{N(0)} s_n^1 z_n^{1*T} y_n^{1*} \right).$$

where the y^{t*} and Z^{t*} are the demeaned y^t and Z^t as in the previous section. Now premultiply both sides of equation (56) by the matrix $[\Sigma_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \Sigma_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}]$ and we obtain the following equation that is the weighted counterpart to equation (40):

$$(71) \qquad \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} \right] \gamma^*$$

$$= \sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} y_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} y_n^{1*}$$

$$= \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} \right] \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} \right]^{-1} \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} y_n^{0*} \right]$$

$$+ \left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} \right] \left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} \right]^{-1} \left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} y_n^{1*} \right]$$

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

using (69) and (70).

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

$$(72) \quad 2\left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}\right]\left[\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*\right]$$

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

$$- 2\left\{\left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*}\right]\boldsymbol{\gamma}^{0*} + \left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}\right]\boldsymbol{\gamma}^{1*}\right\} \quad \text{using (71).}$$

$$= \left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*}\right][\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*}] - \left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}\right][\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*}]$$

$$= -\left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} - \sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*}\right][\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*}]$$

Now premultiply both sides of equation (72) by $(1/2)[\Sigma_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \Sigma_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}]^{-1}$ and substitute the resulting expression for $(1/2)\boldsymbol{\gamma}^{0*} + (1/2)\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*$ 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:

$$(73) \quad \text{LP}_{\text{WHD}} - \text{LP}_{\text{WHI}} = (\mathbf{s}^{1T}\mathbf{Z}^1 - \mathbf{s}^{0T}\mathbf{Z}^0)\left(\frac{1}{2}\boldsymbol{\gamma}^{0*} + \frac{1}{2}\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^*\right)$$

$$= -\frac{1}{2}[\mathbf{s}^{1T}\mathbf{Z}^1 - \mathbf{s}^{0T}\mathbf{Z}^0]\left[\sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*} + \sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}\right]^{-1}$$

$$\left[\sum_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*} - \sum_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*}\right][\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*}].$$

Using equation (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:

- $\mathbf{s}^{1T}\mathbf{Z}^1$ equals $\mathbf{s}^{0T}\mathbf{Z}^0$ so that the period expenditure share weighted amount of each characteristic across models in each period stays the same
- $\Sigma_{n=1}^{N(1)} s_n^1 z_n^{1*T} z_n^{1*}$ equals $\Sigma_{n=1}^{N(0)} s_n^0 z_n^{0*T} z_n^{0*}$ so that the expenditure share weighted model characteristics variance covariance matrix is the same in the two periods
- $\boldsymbol{\gamma}^{1*}$ equals $\boldsymbol{\gamma}^{0*}$ 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 $\mathbf{s}^{1T}\mathbf{Z}^1$ is close to $\mathbf{s}^{0T}\mathbf{Z}^0$, 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 demonstrated in this chapter. If, however, $\mathbf{s}^{1T}\mathbf{Z}^1$ is not close to $\mathbf{s}^{0T}\mathbf{Z}^0$, 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, $\boldsymbol{\gamma}^{0*}$ and $\boldsymbol{\gamma}^{1*}$, then the method of quality adjustment could matter. If equations (57) and (58) are run together as a pooled regression model and an F test rejects the equality of $\boldsymbol{\gamma}^0$ and $\boldsymbol{\gamma}^1$, then it seems sensible to use the weighted hedonic imputation method, which does not depend on having $\boldsymbol{\gamma}^0$ equal to $\boldsymbol{\gamma}^1$ as does the hedonic time dummy method. If the F test does not reject the equality of $\boldsymbol{\gamma}^0$ and $\boldsymbol{\gamma}^1$ 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 $\boldsymbol{\gamma}^0$ and $\boldsymbol{\gamma}^1$ 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 4.7.

4.6   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 nonspecialized PC store-type) representing a sales volume of 1.5 million models worth £1.57 billion. Table 4.1 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).[22] 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 latter, for the hedonic dummy indexes. The estimated

---

22. Bear in mind that some of the indexes estimated in this chapter 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 percent of the January sales value and a mere 12 percent of the December sales value.

Table 4.1   **Number of matched and unmatched observations**

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

Note: Figures are for comparisons between January and each current month.

coefficients for speed in the hedonic regressions were statistically significant coefficients with the expected positive signs.[23]

Table 4.2 columns (1) and (2) show falls in the *unweighted* hedonic dummy and hedonic imputation indexes of 73.2 and 76.4 percent, respectively, a difference of 3.2 percentage points. Columns (4) through (6) show the constituent elements of equation (41) that make up this difference: column (4) is the change in mean characteristics, $[1_1^T \mathbf{Z}^1/N(1)] - [1_0^T \mathbf{Z}^0/N(0)]$, 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, $[\mathbf{Z}^{0*T}\mathbf{Z}^{0*} + \mathbf{Z}^{1*T}\mathbf{Z}^{1*}]^{-1}[\mathbf{Z}^{1*T}\mathbf{Z}^{1*} - \mathbf{Z}^{0*T}\mathbf{Z}^{0*}]$, with one characteristic in this illustration it is the relative change in the variance of speed, falling in some early months but increasing thereafter. Column (6) is the change in the characteristic parameter estimates, $[\gamma^{1*} - \gamma^{0*}]$, the estimated parameters decreased over time. The decomposition of the difference between the hedonic dummy and imputation indexes given in column (3) is exactly 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 *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

23. 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 percent 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.

Table 4.2   **Decomposition of differences between unweighted hedonic time dummy and imputation indexes for desktop PCs, 1998**

	LP_{HD} (1)	LP_{HI} (2)	$LP_{HD}-LP_{HD}$ (3)	Change in mean characteristics (4)	Relative change in var-cov matrices (5)	Change in parameters (6)	$(4)\times(5)\times(6)/2$ (7)
			Change relative to fixed base, January 1998				
February	-0.0837	-0.0837	0.0000	-0.02	0.0564	-0.00042	0.0000
March	-0.1609	-0.1611	0.0003	7.69	0.1485	-0.00045	0.0003
April	-0.2717	-0.2713	-0.0004	47.16	-0.1895	-0.00010	-0.0004
May	-0.4334	-0.4348	0.0014	50.93	-0.1444	0.00039	0.0014
June	-0.4359	-0.4347	-0.0012	60.64	-0.0390	-0.00099	-0.0012
July	-0.4535	-0.4540	0.0004	68.99	0.0119	-0.00103	0.0004
August	-0.4560	-0.4672	0.0112	76.71	0.1484	-0.00196	0.0112
September	-0.4272	-0.4467	0.0195	84.15	0.1787	-0.00259	0.0195
October	-0.5727	-0.5945	0.0218	103.89	0.1906	-0.00220	0.0218
November	-0.6705	-0.7046	0.0341	118.27	0.2317	-0.00249	0.0341
December	-0.7320	-0.7643	0.0323	127.47	0.2050	-0.00247	0.0323

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 4.2 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 4.4 and 4.5 *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 4.3. Again the difference depends on the product of three terms: the change in the expenditure share weighted mean of each characteristics, $[\mathbf{s}^{1T}\mathbf{Z}^1 - \mathbf{s}^{0T}\mathbf{Z}^0]$; the change in the expenditure share weighted characteristics variance-covariance matrix, $[\Sigma_{n=1}^{N(0)}s_n^0 z_n^{0*T} z_n^{0*} + \Sigma_{n=1}^{N(1)}s_n^1 z_n^{1*T} z_n^{1*}]^{-1}$ $\times [\Sigma_{n=1}^{N(1)}s_n^1 z_n^{1*T} z_n^{1*} - \Sigma_{n=1}^{N(0)}s_n^0 z_n^{0*T} z_n^{0*}]$; and the change in the parameters estimates, $[\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*}]$ from the separate (weighted) hedonic regressions in each period. Table 4.3 shows respective falls for the weighted hedonic dummy and weighted hedonic imputation indexes of 43.1 and 50.6 percent (imputation) over the year, compared with falls in the corresponding unweighted indexes of 73.2 and 76.4 percent in table 4.2, thus weighting matters. The decomposition in table 4.3 follows equation (73) and is exact with (one-half of) the product of columns (4) to (6) (in column [7]) equaling the difference between the formulas in column (3). The differences are at their highest in September and December at over 7 percentage points, in part due to the relatively high parameter changes. Again comparing September with December, the parameter changes in column (6) are similarly high at –0.0039. But again the change in the mean speed in column (4) at 139.6 MHz in December is much higher than its September figure of 94.45 MHz.[24] This should translate into a very much lower difference between the formulas in September 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 previous empirical example 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

---

24. 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.

Table 4.3   Decomposition of differences between weighted hedonic time dummy and imputation indexes for desktop PCs, 1998

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

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.[25] 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 in the matrix format of equation (73) and as table 4.4.

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 4.2 and 4.3 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.[26] 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.

25. The adjusted-R^2 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.

26. There are statistical caveats to this statement; that is, 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.

Table 4.4   **Factors contributing to difference between weighted HD and weighted HI estimates for PCs for three variables: January to December comparison**

LP_{HI}	−0.4311	0.0001
LP_{HI}	−0.5057	−0.0004
LP_{HD}–LP_{HD}	0.0745	0.0012

Factors contributing to change

Change in mean characteristics		Change in variances	
Speed	139.593	Speed	1,728.1
CAP	180.140	CAP	−3,672.3
RAM	14.405	RAM	167.3

Inverse of total variances		Change covariances	
Speed	6,245.2	Speed:CAP	5,168.0
CAP	55,658.7	Speed:RAM	460.2
RAM	1,131.1	CAP:RAM	−1,93.9

Inverse of total covariances		Change in parameter estimates	
Speed:CAP	5,422.8	Speed	−0.00743
Speed:RAM	637.2	CAP	0.00127
CAP:RAM	3,626.1	RAM	−0.00612

The four terms in equation (73) are reproduced following as matrices, and as table 4.4, for the December with January comparison, the product of which is equal to the difference between the formulas, that is

$$-\frac{1}{2}[\mathbf{s}^{1T}\mathbf{Z}^1 - \mathbf{s}^{0T}\mathbf{Z}^0]\left[\sum_{n=1}^{N(0)}\mathbf{s}_n^0 z_n^{0*T} z_n^{0*} + \sum_{n=1}^{N(1)}\mathbf{s}_n^1 z_n^{1*T} z_n^{1*}\right]^{-1}$$

$$\left[\sum_{n=1}^{N(1)}\mathbf{s}_n^1 z_n^{1*T} z_n^{1*} - \sum_{n=1}^{N(0)}\mathbf{s}_n^0 z_n^{0*T} z_n^{0*}\right][\gamma^{1*} - \gamma^{0*}]$$

$$= -\frac{1}{2}[139.593, 180.140, 14.405] \times$$

$$\begin{bmatrix} 6,245.2 & 5,422.8 & 637.2 \\ 5,422.8 & 55,658.7 & 3,626.1 \\ 637.2 & 3,626.1 & 1,131.0 \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} 1,728.1 & 5,168.0 & 460.2 \\ 5,168.0 & -3,672.3 & -193.9 \\ 460.2 & -193.9 & 167.3 \end{bmatrix}\begin{bmatrix} -0.00743 \\ 0.00127 \\ -0.00612 \end{bmatrix}$$

$$= 0.1182.$$

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 equation [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.

4.7   Concluding Remarks on the Choice between Hedonic Imputation Indexes and Hedonic Dummy 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 twelve 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 imputa-

tion 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 previous considerations, the advantages and disadvantages of the two methods can be seen: HI is "better" because it allows for changing characteristics prices over time; that is, it is more "flexible" but at the cost of:

- Using up more degrees of freedom.
- 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 favor 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 moti-

vation 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 chapter has provided answers to this question.

Appendix A

An Alternative Interpretation of the Hedonic Imputation Estimate of Log Price Change

In section 4.3, we derived an estimator for the logarithm of overall price change between the two periods, LP_{HI} defined by equation (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); that is, 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, 111–12). 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 4.3. Recall the notation used in section 4.3 where $\mathbf{y}^0$ was the $N(0)$ dimensional vector of log model prices in period 0, $\mathbf{y}^1$ was the $N(1)$ dimensional vector of log model prices in period 1 and $\mathbf{Z}^t$ 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, $\mathbf{y}^{t*}$, and the sample average vectors of model characteristics, $\mathbf{z}^{t*}$, as follows:

$$\text{(A1)} \qquad \mathbf{y}^{0*} \equiv \frac{\mathbf{1}_0^T\mathbf{y}^0}{N(0)}; \mathbf{y}^{1*} \equiv \frac{\mathbf{1}_1^T\mathbf{y}^1}{N(1)}; \mathbf{z}^{0*} \equiv \frac{\mathbf{1}_0^T\mathbf{Z}^0}{N(0)}; \mathbf{z}^{1*} \equiv \frac{\mathbf{1}_1^T\mathbf{Z}^1}{N(1)}.$$

Using definitions (A1), equations (28) and (29) in section 4.3 can be rewritten as follows:

$$\text{(A2)} \qquad\qquad\qquad \mathbf{y}^{0*} = \beta_0^* + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*};$$

$$\text{(A3)} \qquad\qquad\qquad \mathbf{y}^{1*} = \beta_1^* + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*}.$$

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:

$$\text{(A4)} \qquad\qquad \beta_1^* - \beta_0^* = \mathbf{y}^{1*} - \mathbf{y}^{0*} - \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*} + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*}.$$

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, LP_L, and the *Paasche-type measure of log price change,* LP_P, as follows:

$$\text{(A5)} \qquad\qquad LP_L \equiv \beta_1^* + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{1*} - (\beta_0^* + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*});$$

$$\text{(A6)} \qquad\qquad LP_P \equiv \beta_1^* + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*} - (\beta_0^* + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{0*}).$$

It can be seen that equations (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, $\mathbf{z}^{0*}$, while the Paasche-type measure holds the characteristics vector constant at the average of the period 1 levels, $\mathbf{z}^{1*}$. 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, LP_F:

$$\text{(A7)} \quad LP_F \equiv \frac{1}{2}(LP_L + LP_P)$$

$$= \frac{1}{2}[\beta_1^* + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{1*} - (\beta_0^* + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*}) + \beta_1^* + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*}$$
$$\qquad - (\beta_0^* + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{0*})] \qquad\qquad \text{using (A6) and (A7)}$$

$$= \beta_1^* - \beta_0^* + \frac{1}{2}(\mathbf{z}^{0*T} + \mathbf{z}^{1*T})(\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*})$$

$$= \mathbf{y}^{1*} - \mathbf{y}^{0*} - \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*} + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*}$$

$$\qquad + \frac{1}{2}(\mathbf{z}^{0*T}\boldsymbol{\gamma}^{1*} - \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*} + \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*} - \mathbf{z}^{1*T}\boldsymbol{\gamma}^{0*}) \quad \text{using (A4)}$$

$$= \mathbf{y}^{1*} - \mathbf{y}^{0*} + \frac{1}{2}(\mathbf{z}^{0*T}\boldsymbol{\gamma}^{1*} + \mathbf{z}^{0*T}\boldsymbol{\gamma}^{0*} - \mathbf{z}^{1*T}\boldsymbol{\gamma}^{1*} - \mathbf{z}^{1*T}\boldsymbol{\gamma}^{0*})$$

$$= \mathbf{y}^{1*} - \mathbf{y}^{0*} - \frac{1}{2}(\boldsymbol{\gamma}^{1*} - \boldsymbol{\gamma}^{0*})^T(\mathbf{z}^{1*} - \mathbf{z}^{0*})$$

$$= LP_{HI},$$

where LP_{HI} was the hedonic imputation index defined by equation (34) in section 4.2. Thus, the two hedonic imputation methods for defining an estimate of price change coincide in the unweighted case.[27]

Now consider the weighted hedonic imputation model that was described in section 4.5. The equally weighted sample averages of the log prices ($\mathbf{y}^{0*}$ and $\mathbf{y}^{1*}$) and of the model characteristics (the vectors $\mathbf{z}^{0*}$ and $\mathbf{z}^{1*}$) defined in equation (A1) are now replaced by the following *expenditure share weighted averages:*

(A8) $\mathbf{y}^{0*} \equiv \mathbf{s}^{0T}\mathbf{y}^0; \mathbf{y}^{1*} \equiv \mathbf{s}^{1T}\mathbf{y}^1; \mathbf{z}^{0*} \equiv \mathbf{s}^{0T}\mathbf{Z}^0; \mathbf{z}^{1*} \equiv \mathbf{s}^{1T}\mathbf{Z}^1.$

Using the new definitions in equation (A8), it can be seen that equations (61) and (62) in section 4.5 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 equations (A5) and (A4) where we use the new hedonic regression estimates for the period 0 weighted regression, β_0^* and γ^{0*}, and for the period 1 weighted regression, β_1^* and γ^{1*}, and the period 0 weighted average characteristics vector $\mathbf{z}^{0*}$ for the Laspeyres measure LP_L and the period 1 weighted average characteristics vector $\mathbf{z}^{1*}$ 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 equation (A7) and again we can show that this Fisher measure is equal to the weighted hedonic imputation index LP_{WHI} defined by equation (65).

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

Appendix B

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) $\phi_L^* \equiv \delta_1^* - \beta_0^*,$

where δ_1^* and β_0^* are defined by equations (32) and (30), respectively. These last two equations can be rewritten as follows:

(A10) $N(1)\delta_1^* = \mathbf{1}_1^T\mathbf{y}^1 - \mathbf{1}_1^T\mathbf{Z}^1\gamma^{0*};$

(A11) $N(0)\beta_0^* = \mathbf{1}_0^T\mathbf{y}^0 - \mathbf{1}_0^T\mathbf{Z}^0\gamma^{0*}.$

---

27. It can also be verified that LP_L is equal to $\beta_1^* - \delta_0^*$ where δ_0^* was defined by equation (33) and LP_P is equal to $\delta_1^* - \beta_0^*$ where δ_1^* was defined by equation (32).

Recall also that the period 0 hedonic regression was written as equation (26) where β_0^* was the period 0 estimated log price level and $\boldsymbol{\gamma}^{0*}$ 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 $\mathbf{y}^1$. After subtracting these quality adjustments from the vector of period 1 log prices $\mathbf{y}^1$, we are left with the period 1 vector $\mathbf{v}^1$ of quality-adjusted prices, less the period 0 log price level defined as follows:

(A12) $\mathbf{v}^1 \equiv \mathbf{y}^1 - (1_1\beta_0^* + \mathbf{Z}^1\boldsymbol{\gamma}^{0*}) = (\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{0*}) - 1_1\beta_0^*.$

Now run a least squares regression of the period 1 residual vector $\mathbf{v}^1$ on a constant term with coefficient ϕ_0. The resulting least squares estimator for ϕ_0 is:

$$
\begin{aligned}
\text{(A13)} \qquad \phi_0^* &\equiv \frac{1_1^T\mathbf{v}^1}{N(1)} \\[2mm]
&= \frac{1_1^T(\mathbf{y}^1 - \mathbf{Z}^1\boldsymbol{\gamma}^{0*})}{N(1)} - \frac{N(1)\beta_0^*}{N(1)} \qquad \text{using (A12)} \\[2mm]
&= \delta_1^* - \beta_0^* \qquad\qquad\qquad\qquad\ \text{using (A10)} \\[2mm]
&= \phi_L^* \qquad\qquad\qquad\qquad\qquad\ \text{using (A9).}
\end{aligned}
$$

Thus, the Laspeyres type hedonic imputation measure of log price change ϕ_L^* defined by equation (A9) is numerically equal to the least squares estimator ϕ_0^* of the constant term in a regression of period 1 quality-adjusted log prices $\mathbf{v}^1$ defined by equation (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.[28]

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

(A14) $\phi_P^* \equiv \beta_1^* - \delta_0^*,$

---

28. This is only an approximate standard error because it is conditional on period 0 estimated parameters, β_0^* and $\boldsymbol{\gamma}_0^*$, 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, β_0, and the constant quality period 1 log price level, δ_1, then we would use the period 0 estimated characteristics prices $\boldsymbol{\gamma}^{0*}$ in order to form period 0 and 1 quality-adjusted log price vectors $\mathbf{w}^0$ and $\mathbf{w}^1$, defined as $\mathbf{w}^t \equiv \mathbf{y}^t - \mathbf{Z}^t\boldsymbol{\gamma}^{0*}$ for $t = 0,1$. Now form two auxiliary regressions where $\mathbf{w}^0$ is regressed on a constant with coefficient β_0 and $\mathbf{w}^1$ is regressed on a constant with coefficient δ_1. The least squares estimators for β_0 and δ_1 turn out to be the β_0^* and δ_1^*, defined by equations (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, $\boldsymbol{\gamma}^{0*}$. 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 β_0.

where β_1^* and δ_0^* are defined by equations (31) and (33), respectively. These last two equations can be rewritten as follows:

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

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

Recall also that the period 0 hedonic regression was written as equation (27), where β_1^* was the period 1 estimated log price level and $\boldsymbol{\gamma}^{1*}$ 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 $\mathbf{y}^0$. After subtracting these quality adjustments from the vector of period 0 log prices $\mathbf{y}^0$, we are left with the period 0 vector $\mathbf{v}^0$ of quality-adjusted prices, less the period 1 log price level defined as follows:

(A17) $$\mathbf{v}^0 \equiv \mathbf{y}^0 - [1_0\beta_1^* + \mathbf{Z}^0\boldsymbol{\gamma}^{1*}] = [\mathbf{y}^0 - \mathbf{Z}^0\boldsymbol{\gamma}^{1*}] - 1_0\beta_1^*.$$

Now run a least squares regression of the period 0 residual vector $\mathbf{v}^0$ on a constant term with coefficient ϕ_1. The resulting least squares estimator for ϕ_1 is:

(A18) $$\phi_1^* \equiv \frac{1_0^T\mathbf{v}^0}{N(0)}$$

$$= \frac{1_0^T(\mathbf{y}^0 - Z^0\boldsymbol{\gamma}^{1*})}{N(0)} - \frac{N(0)\beta_1^*}{N(0)} \qquad \text{using (A17)}$$

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

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

Thus, the Paasche type hedonic imputation measure of log price change ϕ_P^* defined by equation (A14) is numerically equal to minus the least squares estimator ϕ_1^* of the constant term in a regression of period 0 quality-adjusted log prices $\mathbf{v}^0$ defined by equation (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.

References

Berndt, E. R., Z. Griliches, and N. J. Rappaport. 1995. Econometric estimates of price indexes for personal computers in the 1990s. *Journal of Econometrics* 68: 243–68.

Berndt, E. R., and N. J. Rappaport. 2001. Price and quality of desktop and mobile personal computers: A quarter-century historical overview. *American Economic Review* 91 (2): 268–73.

Boskin, M. S. (Chair), Advisory Commission to Study the Consumer Price Index. 1996. Towards a more accurate measure of the cost of living. Interim report to the senate finance committee, Washington DC

Cole, R., Y. C. Chen, J. A. Barquin-Stolleman, E. Dullberger, N. Helvacian, and J. H. Hodge. 1986. Quality-adjusted price indexes for computer processors and selected peripheral equipment. *Survey of Current Business* 66 (1): 41–50.

Court, A. T. 1939. Hedonic price indexes with automotive examples. In *The dynamics of automobile demand,* 98–117. New York: General Motors Corporation.

de Haan, J. 2003. Time dummy approaches to hedonic price measurement. Paper presented at the Seventh Meeting of the International Working Group on Price Indices, (Ottawa Group), 27–29 May, National Institute for Statistics and Economic Studies (INSEE), Paris. Available at: http://www.insee.fr/en/nom_def_met/colloques/ottawa/ottawa_papers.htm.

———. 2004. Hedonic regressions: The time dummy index as a special case of the Törnqvist Index, time dummy approaches to hedonic price measurement. Paper presented at the Eighth Meeting of the International Working Group on Price Indices. 23–25 August, Helsinki: Statistics Finland.

———. 2007. Hedonic price indexes: A comparison of imputation, time dummy and other approaches. Room document at the Tenth Ottawa Group Meeting. 9–12 October, Ottawa. Available at: http://www.ottawagroup2007.ca/r004/pdf/ogo04_033_e.pdf.

Diewert, W. E. 2003a. Hedonic regressions: A consumer theory approach. In *Scanner data and price indexes,* studies in income and wealth, vol. 64, ed. R. C. Feenstra and M. D. Shapiro, 317–48. Chicago: University of Chicago Press.

———. 2003b. Hedonic regressions: A review of some unresolved issues. Paper presented at the Seventh Meeting of the Ottawa Group. 27–29 May, Paris.

———. 2005. Weighted country product dummy variable regressions and index number formulae. *The Review of Income and Wealth* 51 (4): 561–71.

———. 2006. Adjacent period dummy variable hedonic regressions and bilateral index number theory. *Annales d'économie et de statistique* 79/80: 1–28. Available at: http://www.econ.ubc.ca/discpapers/dp0511.pdf.

Dhrymes, P. J. 1971. Price and quality changes in consumer capital goods: An empirical study. In *Price indexes and quality change,* ed. Z. Griliches, 88–149. Cambridge, MA: Harvard University Press.

Feenstra, R. C. 1995. Exact hedonic price indexes. *Review of Economics and Statistics* 77 (4): 634–54.

Griliches, Z. 1967. Hedonic price indexes revisited: A note on the state of the art. *Proceedings of the Business and Economics Section of the American Statistical Association:* 332–34.

———. 1971a. Hedonic price indexes for automobiles: An econometric analysis of quality change. In *Price indexes and quality change,* ed. Z. Griliches, 55–87. Cambridge, MA: Harvard University Press.

———. 1971b. Introduction: Hedonic price indexes revisited. In *Price indexes and quality change,* ed. Z. Griliches, 3–15. Cambridge, MA: Harvard University Press.

Ioannidis, C., and M. Silver. 1999. Estimating exact hedonic indexes: An application to U.K. television sets. *Journal of Economics* 69 (1): 71–94.

Koskimäki, T., and Y. Vartia. 2001. Beyond matched pairs and Griliches-type hedonic methods for controlling quality changes in CPI sub-indices. In *International working group on price indices: Papers and proceedings of the sixth meeting* (the

Ottawa Group) held at Canberra, Australia, April 2–6, ed. K. Woolford, 12–40. Canberra: Australian Bureau of Statistics.

Pakes, A. 2003. A reconsideration of hedonic price indexes with an application to PCs. *The American Economic Review* 93 (5): 1576–93.

Schultze, C. L., and C. Mackie, eds. 2002. *At what price? Conceptualizing and measuring cost-of-living indexes,* Washington DC: National Academy Press.

Silver, M., and S. Heravi. 2003. The measurement of quality-adjusted price changes. In *Scanner data and price indexes,* ed. M. Shapiro and R. Feenstra, 277–317. Chicago: University of Chicago Press.

———. 2005. A failure in the measurement of inflation: Results from a hedonic and matched experiment using scanner data. *Journal of Business and Economic Statistics* 23 (5): 269–81.

———. 2007a. Hedonic imputation indexes and time dummy hedonic indexes. *Journal of Business and Economic Statistics* 25 (2): 239–46.

———. 2007b. Different approaches to estimating hedonic indexes. In *Hard-to-measure goods and services: Essays in honour of Zvi Griliches,* ed. E. R. Berndt and C. Hulten, 235–68. Chicago: University of Chicago Press.

Stigler, G., ed. 1961. *The price statistics of the federal government,* National Bureau of Economic Research. Chicago: University of Chicago Press.

Triplett, J. 2004. *Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products.* Directorate for Science, Technology and Industry Working Paper 2004/9, Paris: Organization for Economic Cooperation and Development (OECD).

Triplett, J. E., and R. J. McDonald. 1977. Assessing the quality error in output measures: The case of refrigerators. *The Review of Income and Wealth* 23 (2): 137–56.

Comment   Jan de Haan

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 chapter seems highly relevant for the work of statistical agencies. However, the use of matrix algebra makes the exposition very

Jan de Haan is a senior researcher at Statistics Netherlands.