Hedonic Regressions: A Review of Some Unresolved Issues

Erwin Diewert

University of British Columbia, Vancouver, Canada

The author is indebted to Ernst Berndt and Alice Nakamura for helpful comments.

1. Introduction

Three recent publications have revived interest in the topic of hedonic regressions. The first publication is Pakes (2001) who proposed a somewhat controversial view of the topic.[1] The second publication is Chapter 4 in Schultze and Mackie (2002), where a rather cautious approach to the use of hedonic regressions was advocated due to the fact that many issues had not yet been completely resolved. A third paper by Heravi and Silver (2002) also raised questions about the usefulness of hedonic regressions since this paper presented several alternative hedonic regression methodologies and obtained different empirical results using the alternative models.[2]

Some of the more important issues that need to be resolved before hedonic regressions can be routinely applied by statistical agencies include:

- Should the dependent variable be transformed or not?
- Should separate hedonic regressions be run for each of the comparison periods or should we use the dummy variable adjacent year regression technique initially suggested by Court (1939; 109-11) and used by Berndt, Griliches and Rappaport (1995; 260) and many others?
- Should regression coefficients be sign restricted or not?
- Should the hedonic regressions be weighted or unweighted? If they should be weighted, should quantity or expenditure weights be used?[3]
- How should outliers in the regressions be treated? Can influence analysis be used?

The present paper takes a systematic look at the above questions. Single period hedonic regression issues are addressed in sections 2 to 5 while two year time dummy variable regression issues are addressed in sections 6 and 7. Some of the more technical material relating to section 7 is in an Appendix, which examines the properties of bilateral weighted

[1] See Hulten (2002) for a nice review of the issues raised in Pakes paper.

[2] The observation that different variants of hedonic regression techniques can generate quite different answers empirically dates back to Triplett and McDonald (1977; 150) at least.

[3] Diewert (2002b) recently looked at these weighting issues in the context of a simplified adjacent year hedonic regression model where the only characteristics were dummy variables.

hedonic regressions. Section 8 discusses the treatment of outliers and influential observations and section 9 addresses the issue of whether the signs of hedonic regression coefficients should be restricted. Section 10 concludes.

2. To Log or Not to Log

We suppose that price data have been collected on K models or varieties of a commodity over T+1 periods.[4] Thus p_k^t is the price of model k in period t for t = 0,1,...,T and $k \in S(t)$ where S(t) is the set of models that are actually sold in period t. For $k \in S(t)$, denote the number of these type k models sold during period t by q_k^t.[5] We suppose also that information is available on N relevant characteristics of each model. The amount of characteristic n that model k possesses in period t is denoted as z_{kn}^t for t = 0,1,...,T, n = 1,...,N and $k \in S(t)$. Define the N dimensional vector of characteristics for model k in period t as $z_k^t \equiv [z_{k1}^t, z_{k2}^t, ..., z_{kN}^t]$ for t = 0,1,...,T and $k \in S(t)$. We shall consider only linear hedonic regressions in this review. Hence, the *unweighted linear hedonic regression for period t* has the following form:[6]

$$(1)\quad f(p_k^t) = \beta_0^t + \sum_{n=1}^{N} f_n(z_{kn}^t)\beta_n^t + \varepsilon_k^t ; \qquad t = 0,1,...,T;\ k \in S(t)$$

where ε_k^t is an independently distributed error term with mean 0 and variance σ^2, f(x) is either the identity function $f(x) \equiv x$ or the natural logarithm function $f(x) \equiv \ln x$ and the functions of one variable f_n are either the identity function, the logarithm function or a dummy variable which takes on the value 1 if the characteristic n is present in model k or 0 otherwise. We are restricting the f and f_n in this way since the identity, log and dummy variable functions are by far the most commonly used transformation functions used in hedonic regressions.

Recall that the period t characteristics vector for model k was defined as $z_k^t \equiv [z_{k1}^t, z_{k2}^t, ..., z_{kN}^t]$. We define also the period t vector of the β's as $\beta^t \equiv [\beta_0^t, \beta_1^t, ..., \beta_N^t]$. Using these definitions, we simplify the notation on the right hand side of (1) by defining:

$$(2)\quad h^t(z_k^t, \beta^t) \equiv \beta_0^t + \sum_{n=1}^{N} f_n(z_{kn}^t)\beta_n^t \qquad t = 0,1,...,T;\ k \in S(t).$$

The question we now want to address is: should the dependent variable $f(p_k^t)$ on the left hand side of (1) be p_k^t or $\ln p_k^t$; i.e., should f be the identity function or the log function?[7] We also would like to know if the choice of identity or log for the function f should affect our choice of identity or log for the f_n that correspond to the continuous (i.e., non dummy variable) characteristics.

Suppose that we choose f to be the identity function. Suppose further that there is only one continuous characteristic so that N = 1. In this situation, the hedonic regression is essentially a regression of price on package size and so if we want to have as a special case, that price per

[4] Models sold in different outlets can be regarded as separate varieties or not, depending on the context.
[5] If a particular model k is sold at various prices during period t, then we interpret q_k^t as the total quantity of model k that is sold in period t and p_k^t as the corresponding average price or unit value.
[6] Note that the linear regression model defined by (1) can only provide a first order approximation to a general hedonic function. Diewert (2001) made a case for considering second order approximations but in this paper, we will follow current practice and consider only linear approximations.
[7] Griliches (1971a; 58) noted that an advantage of the log formulation is that β_n^t would provide an estimate of the percentage change in price due to a one unit change in z_n, provided that f_n was the identity function. Court (1939; 111) implicitly noted this advantage of the log formulation.

unit of useful characteristic is a constant, then we should set $f_1(z_1) = z_1$.[8] Under these conditions, the model defined by (1) and f(p) = p will be consistent with the constant per unit price hypothesis if $\beta_0^t = 0$. In the case of N continuous characteristics, a generalization of the constant per unit characteristic price hypothesis is the hypothesis of constant returns to scale in the vector of characteristics, so that if all characteristics are doubled, then the resulting model price is doubled. If our period t model is defined by (1) and f(p) = p, then h^t must satisfy the following property:

$$(3)\quad \beta_0^t + \sum_{n=1}^N f_n(\lambda z_{kn}^t)\beta_n^t = \lambda[\beta_0^t + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n^t] \qquad \text{for all } \lambda > 0.$$

In order to satisfy (3), we must choose $\beta_0^t = 0$ and the f_n to be identity functions. Thus if f is chosen to be the identity function, then it is natural to choose the f_n that correspond to continuous characteristics to be identity functions as well.[9]

Now suppose that we choose f to be the log function. Suppose again that there is only one continuous characteristic so that N = 1. In this situation, again the hedonic regression is essentially a regression of price on package size and so if we want to have as a special case, that price per unit of useful characteristic is a constant, then we need to set $f_1(z_1) = \ln z_1$ and $\beta_1^t = 1$. Under these conditions, the model defined by (1) and f(p) = lnp will be consistent with the constant per unit price hypothesis. In the case of N continuous characteristics, a generalization of the constant per unit price hypothesis is the hypothesis of constant returns to scale in the vector of characteristics. If our period t model is defined by (1) and f(p) = lnp, then h^t must satisfy the following property:

$$(4)\quad \beta_0^t + \sum_{n=1}^N f_n(\lambda z_{kn}^t)\beta_n^t = \ln\lambda + \beta_0^t + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n^t \qquad \text{for all } \lambda > 0.$$

In order to satisfy (4), we must choose the $f_n(z_n)$ to be log functions[10] and the β_n^t must satisfy the following linear restriction:

$$(5)\quad \sum_{n=1}^N \beta_n^t = 1.$$

Thus if f is chosen to be the log function, then it is natural to choose the f_n that correspond to continuous characteristics to be log functions as well.

An extremely important property that a hedonic regression model should possess is that the model be invariant to changes in the units of measurement of the continuous characteristics. Thus suppose that we have only continuous characteristics and the period t model is defined by (1) with f arbitrary and the $f_n(z_n) = \ln z_n$. Suppose further that new units of measurement for the N characteristics are chosen, say Z_n, where

$$(6)\quad Z_n \equiv z_n/c_n ; \qquad n = 1,...,N$$

[8] We are not arguing that this constant returns to scale hypothesis must necessarily hold (usually, it will not hold); we are just arguing that it is useful for the hedonic regression model to be able to model this situation as a special case. The constant returns to scale hypothesis is required in some hedonic models; e.g., see Muellbauer (1974; 988) and Pollak's (1983) "L Characteristics" model, which is also used by Triplett (1983).

[9] If we change the units of measurement for the continuous characteristics, then the linear hedonic regression model will be unaffected by this change in the units; i.e., the change in the units for the nth characteristic can be absorbed into the regression coefficient β_n.

[10] Note that all of the continuous characteristics must be measured in positive units in this case.

where the c_n are positive constants. The invariance property requires that we can find new regression coefficients, β_n^{t*}, such that the following equation can be satisfied identically:

$$\begin{aligned}(7)\ \beta_0^t + \sum_{n=1}^N (\ln z_n)\beta_n^t &= \beta_0^{t*} + \sum_{n=1}^N (\ln Z_n)\beta_n^{t*} \\ &= \beta_0^{t*} + \sum_{n=1}^N (\ln z_n/c_n)\beta_n^{t*} \qquad \text{using (6)} \\ &= \beta_0^{t*} - \sum_{n=1}^N (\ln c_n)\beta_n^{t*} + \sum_{n=1}^N (\ln z_n)\beta_n^{t*}.\end{aligned}$$

Hence to satisfy (7) identically, we need only set $\beta_n^{t*} = \beta_n^t$ for n = 1,...,N and set $\beta_0^{t*} = \beta_0^t - \sum_{n=1}^N (\ln c_n)\beta_n^t$. Thus in particular, the hedonic regression model where f and the f_n are all log functions will satisfy the important invariance to changes in the units of measurement of the continuous characteristics property, provided that the regression has a constant term in it.[11]

We now address the following question: should the dependent variable $f(p_k^t)$ on the left hand side of (1) be p_k^t or $\ln p_k^t$?

If f is the identity function, then using definitions (2), equations (1) can be rewritten as follows:

$$(8)\ p_k^t = h^t(z_k^t,\beta^t) + \varepsilon_k^t ; \qquad t = 0,1,...,T;\ k \in S(t)$$

where ε_k^t is an independently distributed error term with mean 0 and variance σ^2. On the other hand, if f is the logarithm function, then equations (1) are equivalent to the following equations:

$$\begin{aligned}(9)\ p_k^t &= \exp[h^t(z_k^t,\beta^t)]\exp[\varepsilon_k^t] ; \qquad t = 0,1,...,T;\ k \in S(t) \\ &= \exp[h^t(z_k^t,\beta^t)]\eta_k^t ;\end{aligned}$$

where η_k^t is an independently distributed error term with mean 1 and constant variance. Which is more plausible: the model specified by (8) or the model specified by (9)? We argue that it is more likely that the errors in (9) are homoskedastic compared to the errors in (8) since models with very large characteristic vectors z_k^t will have high prices p_k^t and are very likely to have relatively large error terms. On the other hand, models with very small amounts of characteristics will have small prices and small means and the deviation of a model price from its mean will be necessarily small. In other words, it is more plausible to assume that the ratio of model price to its mean price is randomly distributed with mean 1 and constant variance than to assume that the difference between model price and its mean is randomly distributed with mean 0 and constant variance. Hence, from an a priori point of view, we would favor the logarithmic regression model (9) (or (1) with $f(p) \equiv \ln p$) over its linear counterpart (8).

The regression models considered in this section were unweighted models and could be estimated without a knowledge of the amounts sold for each model in each period. In the following section, we assume that model quantity information q_k^t is available and we consider how this extra information could be used.

[11] Note that the above argument is independent of the functional form for f; i.e., if the f_n for the continuous characteristics are log functions, then for any f, the hedonic regression must include a constant term to be invariant to changes in the units of these continuous characteristics.

3. Quantity Weights versus Expenditure Weights

Usually, discussions of how to use quantity or expenditure weights in a hedonic regression are centered around discussions on how to reduce the heteroskedasticity of error terms. In this section, we attempt a somewhat different approach based on the idea that the regression model should be *representative*. In other words, if model k sold q_k^t times in period t, then perhaps model k should be repeated in the period t hedonic regression q_k^t times so that the period t regression is representative of the sales that actually occurred during the period.[12]

To illustrate this idea, suppose that in period t, only three models were sold and there is only one continuous characteristic. Let the period t price of the three models be p_1^t, p_2^t and p_3^t and suppose that the three models have the amounts z_{11}^t, z_{21}^t and z_{31}^t of the single characteristic respectively. Then the period t unweighted regression model (1) has only the following 3 observations and 2 unknown parameters, β_0^t and β_1^t :

$$(10)\quad f(p_1^t) = \beta_0^t + f_1(z_{11}^t)\beta_1^t + \varepsilon_1^t ;$$
$$f(p_2^t) = \beta_0^t + f_1(z_{21}^t)\beta_1^t + \varepsilon_2^t ;$$
$$f(p_3^t) = \beta_0^t + f_1(z_{31}^t)\beta_1^t + \varepsilon_3^t .$$

Note that each of the 3 observations gets an equal weight in the period t hedonic regression model defined by (10). However, if say models 1 and 2 are vastly more popular than model 3, then it does not seem to be appropriate that model 3 gets the same importance as models 1 and 2.

Suppose that the integers q_1^t, q_2^t and q_3^t are the amounts sold in period t of models 1,2 and 3 respectively. Then one way of constructing a hedonic regression that weights models according to their economic importance is to *repeat* each model observation according to the number of times it sold in the period. This leads to the following more representative hedonic regression model, where the error terms have been omitted:

$$(11)\quad 1_1 f(p_1^t) = 1_1\beta_0^t + 1_1 f_1(z_{11}^t)\beta_1^t ;$$
$$1_2 f(p_2^t) = 1_2\beta_0^t + 1_2 f_1(z_{21}^t)\beta_1^t ;$$
$$1_3 f(p_3^t) = 1_3\beta_0^t + 1_3 f_1(z_{31}^t)\beta_1^t$$

where 1_k is a vector of ones of dimension q_k^t for k = 1,2,3.

Now consider the following quantity transformation of the original unweighted hedonic regression model (10):

$$(12)\quad (q_1^t)^{1/2} f(p_1^t) = (q_1^t)^{1/2} \beta_0^t + (q_1^t)^{1/2} f_1(z_{11}^t)\beta_1^t + \varepsilon_1^{t*} ;$$
$$(q_2^t)^{1/2} f(p_2^t) = (q_2^t)^{1/2} \beta_0^t + (q_2^t)^{1/2} f_1(z_{21}^t)\beta_1^t + \varepsilon_2^{t*} ;$$
$$(q_3^t)^{1/2} f(p_3^t) = (q_3^t)^{1/2} \beta_0^t + (q_3^t)^{1/2} f_1(z_{31}^t)\beta_1^t + \varepsilon_3^{t*} .$$

[12] Thus our representative approach follows along the lines of Theil's (1967; 136-138) stochastic approach to index number theory, which is also pursued by Rao (2002). The use of weights that reflect the economic importance of models was recommended by Griliches (1971b; 8): "But even here, we should use a weighted regression approach, since we are interested in an estimate of a weighted average of the pure price change, rather than just an unweighted average over all possible models, no matter how peculiar or rare." However, he did not make any explicit weighting suggestions.

Comparing (10) and (12), it can be seen that the observations in (12) are equal to the corresponding observations in (10), except that the dependent and independent variables in observation k of (10) have been multiplied by the square root of the quantity sold of model k in period t for k = 1,2,3 in order to obtain the observations in (12). A sampling framework for (12) is available if we assume that the transformed residuals ε_k^{t*} are independently normally distributed with mean zero and constant variance.

Let b_0^t and b_1^t denote the least squares estimators for the parameters β_0^t and β_1^t in (11) and let b_0^{t*} and b_1^{t*} denote the least squares estimators for the parameters β_0^t and β_1^t in (12). Then it is straightforward to show that these two sets of least squares estimators are the same[13]; i.e., we have:

(13) $[b_0^t, b_1^t] = [b_0^{t*}, b_1^{t*}]$.

Thus a shortcut method for obtaining the least squares estimators for the unknown parameters, β_0^t and β_1^t, which occur in the "representative" model (11) is to obtain the least squares estimators for the transformed model (12). This equivalence between the two models provides a justification for using the weighted model (12) in place of the original model (10). The advantage in using the transformed model (12) over the "representative" model (11) is that we can develop a sampling framework for (12) but not for (11), since the (omitted) error terms in (11) cannot be assumed to be distributed independently of each other. However, in view of the equivalence between the least squares estimators for models (11) and (12), we can now be comfortable that the regression model (12) weights observations according to their quantitative importance in period t. Hence, we definitely recommend the use of the weighted hedonic regression model (12) over its unweighted counterpart (10).

However, rather than weighting models by their *quantity* sold in each period, it is possible to weight each model according to the *value* of its sales in each period. Thus define the value of sales of model k in period t to be:

(14) $v_k^t \equiv p_k^t q_k^t$; $\quad t = 0,1,...,T$; $k \in S(t)$.

Now consider again the simple unweighted hedonic regression model defined by (10) above and round off the sales of each of the 3 models to the nearest dollar (or penny). Let 1_{k*} be a vector of ones of dimension v_k^t for k = 1,2,3. Repeating each model in (10) according to the value of its sales in period t leads to the following more representative period t hedonic regression model (where the errors have been omitted):

(15) $1_{1*} f(p_1^t) = 1_{1*}\beta_0^t + 1_{1*} f_1(z_{11}^t)\beta_1^t$;
$1_{2*} f(p_2^t) = 1_{2*}\beta_0^t + 1_{2*} f_1(z_{21}^t)\beta_1^t$;
$1_{3*} f(p_3^t) = 1_{3*}\beta_0^t + 1_{3*} f_1(z_{31}^t)\beta_1^t$.

[13] See, for example, Greene (1993; 277-279). However, the numerical equivalence of the least squares estimates obtained by repeating multiple observations or by the square root of the weight transformation was noticed long ago as the following quotation indicates: "It is evident that an observation of weight w enters into the equations exactly as if it were w separate observations each of weight unity. The best practical method of accounting for the weight is, however, to prepare the equations of condition by multiplying each equation throughout by the square root of its weight." E. T. Whittaker and G. Robinson (1940; 224).

Now consider the following value transformation of the original unweighted hedonic regression model (10):

$$(16)\ (v_1^t)^{1/2} f(p_1^t) = (v_1^t)^{1/2} \beta_0^t + (v_1^t)^{1/2} f_1(z_{11}^t)\beta_1^t + \varepsilon_1^{t**};$$
$$(v_2^t)^{1/2} f(p_2^t) = (v_2^t)^{1/2} \beta_0^t + (v_2^t)^{1/2} f_1(z_{21}^t)\beta_1^t + \varepsilon_2^{t**};$$
$$(v_3^t)^{1/2} f(p_3^t) = (v_3^t)^{1/2} \beta_0^t + (v_3^t)^{1/2} f_1(z_{31}^t)\beta_1^t + \varepsilon_3^{t**}.$$

Comparing (10) and (16), it can be seen that the observations in (12) are equal to the corresponding observations in (10), except that the dependent and independent variables in observation k of (10) have been multiplied by the square root of the value sold of model k in period t for k = 1,2,3 in order to obtain the observations in (16). Again, a sampling framework for (16) is available if we assume that the transformed residuals ε_k^{t**} are independently distributed normal random variables with mean zero and constant variance.

Again, it is straightforward to show that the least squares estimators for the parameters β_0^t and β_1^t in (15) and (16) are the same. Thus a shortcut method for obtaining the least squares estimators for the unknown parameters, β_0^t and β_1^t, which occur in the value weights representative model (15) is to obtain the least squares estimators for the transformed model (16). This equivalence between the two models provides a justification for using the value weighted model (16) in place of the original model (10). As before, the advantage in using the transformed model (16) over the value weights representative model (15) is that we can develop a sampling framework for (16) but not for (15), since the (omitted) error terms in (15) cannot be assumed to be distributed independently of each other.

It seems to us that the quantity weighted and value weighted models are clear improvements over the original unweighted model (10). Our reasoning here is similar to that used by Fisher (1922; Chapter III) in developing bilateral index number theory, who argued that prices needed to be weighted according to their quantitative or value importance in the two periods being compared.[14] In the present context, we have a weighting problem that involves only one period so that our weighting problems are actually much simpler than those considered by Fisher: we need only choose between quantity or value weights!

But which system of weighting is better in our present context: quantity or value weighting?

The problem with quantity weighting is this: it will tend to give too little weight to models that have high prices and too much weight to cheap models that have low amounts of useful characteristics. Hence it appears to us that value weighting is clearly preferable. Thus we are taking the point of view that the main purpose of the period t hedonic regression is to enable

[14] "It has already been observed that the purpose of any index number is to strike a 'fair average' of the price movements—or movements of other groups of magnitudes. At first a *simple* average seemed fair, just because it treated all terms alike. And, in the absence of any knowledge of the relative importance of the various commodities included in the average, the simple average *is* fair. But it was early recognized that there are enormous differences in importance. Everyone knows that pork is more important than coffee and wheat than quinine. Thus the quest for fairness led to the introduction of weighting." Irving Fisher (1922; 43). "But on what principle shall we weight the terms? Arthur Young's guess and other guesses at weighting represent, consciously or un consciously, the idea that relative *money values* of the various commodities should determine their weights. A value is, of course, the product of a price per unit, multiplied by the number of units taken. Such values afford the only common measure for comparing the streams of commodities produced, exchanged, or consumed, and afford almost the only basis of weighting which has ever been seriously proposed." Irving Fisher (1922; 45).

us to decompose the market value of each model sold, $p_k^t q_k^t$, into the product of a period t price for a quality adjusted unit of the hedonic commodity, say P^t, times a constant utility total quantity for model k, Q_k^t. Hence observation k in period t should have the representative weight Q_k^t in constant utility units that are comparable across models. But Q_k^t is equal to $p_k^t q_k^t / P^t$, which in turn is equal to v_k^t / P^t, which in turn is proportional to v_k^t. Thus weighting by the values v_k^t seems to be the most appropriate form of weighting.

Our conclusions about single period hedonic regressions at this point can be summarized as follows:

- With respect to taking transformations of the dependent variable in a period t hedonic regression, taking of logarithms of the model prices is our preferred transformation.

- If information on the number of models sold in each period is available, then weighting each observation by the square root of the value of model sales is our preferred method of weighting.

- If the log transformation is chosen for the dependent variable, then we have a mild preference for transforming the continuous characteristics by the logarithm transformation as well. If the continuous characteristics are transformed by the logarithmic transformation, then the regression must have a constant term to ensure that the results of the regression are invariant to the choice of units for the characteristics.

- If the dependent variable is simply the model price, then we have a mild preference for not transforming the continuous characteristics as well.

With the above general considerations in mind, we now turn to a discussion of how single period hedonic regressions can be used by statistical agencies in a sampling context.

4. The Use of Single Period Hedonic Regressions in a Replacement Sampling Context

In this section, we consider the use of single period hedonic regressions in the context of statistical agency sampling procedures where a sampled model that was available in period s is not available in a later period t and is replaced with a new model that is available in period t.

We assume that $s < t$ and that model 1 is available in period s (with price p_1^s and characteristics vector z_1^s) but is not available in period t. We further assume that model 1 is *replaced* by model 2 in period t, with price p_2^t and characteristics vector z_2^t. The problem is to somehow adjust the price relative p_2^t / p_1^s so that the *adjusted price relative* can be averaged with other price relatives of the form p_k^t / p_k^s that correspond to models k that are present in both periods s and t in order to form an overall price relative for the item level, going from period s to t. If the item level index is a chain type index, then s will be equal to t−1 and if the item level index is a fixed base type index, then s will be equal to the base period 0.

Recall the family of single period hedonic regressions defined in section 2 above by equations (1). If we use definitions (2) and assume that the function of one variable f(x) has an inverse function f^{-1}, then we may rewrite equations (1) as follows:

$$(17)\ p_k^t = f^{-1}[h^t(z_k^t, \beta^t) + \varepsilon_k^t]\ ; \qquad t = 0,1,...,T;\ k \in S(t).$$

Assume that we have a vector of estimates b^t for the period t vector of parameters β^t and define the *model k sample residuals for period t*, e_k^t, as follows:[15]

$$(18)\ e_k^t \equiv f(p_k^t) - h^t(z_k^t,\beta^t)\ ; \qquad t = 0,1,...,T;\ k \in S(t).$$

Thus the sample counterparts to equations (17) are the following equations:

$$(19)\ p_k^t = f^{-1}[h^t(z_k^t,b^t) + e_k^t]\ ; \qquad t = 0,1,...,T;\ k \in S(t).$$

Now suppose that the period s hedonic regression is available to the statistical agency. Thus equation (19) for period s and model 1 is:

$$(20)\ p_1^s = f^{-1}[h^s(z_1^s,b^s) + e_1^s].$$

Recall that model 2, the replacement for model 1 in period t, has the vector of characteristics z_2^t. Hence, using the period s hedonic regression, a comparable price for model 2 in period s is $f^{-1}[h^s(z_2^t,b^s)]$, the predicted period s price using the period t hedonic regression for a model with the vector of characteristics z_2^t. Thus our first estimator for an adjusted price relative for models 1 and 2 going from period s to t is:

$$(21)\ r(1) \equiv p_2^t/f^{-1}[h^s(z_2^t,b^s)].$$

However, there is a problem with the use of (21) as an adjusted price relative. The problem will become apparent if $z_2^t = z_1^s$, so that the two models are in fact identical. In this case, we want our price relative to equal the actual price ratio:

$$(22)\ p_2^t/p_1^s = p_2^t/f^{-1}[h^s(z_1^s,b^s) + e_1^s] \qquad \text{using (20)}$$
$$\neq p_2^t/f^{-1}[h^s(z_1^s,b^s)] \qquad \text{if } e_1^s \neq 0.$$

Hence if the regression residual for model 1 in period s, e_1^s, is not equal to zero, then r(1) defined by (21) will not be an appropriate adjusted price relative. In order to compare like with like, we must multiply r(1) by an adjustment factor equal to

$$(23)\ f^{-1}[h^s(z_1^s,b^s)]/p_1^s = f^{-1}[h^s(z_1^s,b^s)]/f^{-1}[h^s(z_1^s,b^s) + e_1^s].$$

Thus our second estimator r(2) for an adjusted price relative is r(1) defined by (21) times the adjustment factor defined by (23), which adjusts the period s observed price for model 1, p_1^s, onto the period s hedonic regression surface:[16]

$$(24)\ r(2) \equiv \{p_2^t/f^{-1}[h^s(z_2^t,b^s)]\}\{f^{-1}[h^s(z_1^s,b^s)]/p_1^s\}$$
$$= \{p_2^t/f^{-1}[h^s(z_2^t,b^s)]\}/\{p_1^s/f^{-1}[h^s(z_1^s,b^s)]\}.$$

The second expression for r(2) in (24) is instructive. We can interpret $p_2^t/f^{-1}[h^s(z_2^t,b^s)]$ as the period t price for model 2 expressed in constant quality utility units, using the period s hedonic regression as the quality adjustment mechanism. Similarly, we can interpret $p_1^s/f^{-1}[h^s(z_1^s,b^s)]$ as the period s price for model 1 expressed in constant quality utility units,

[15] Definitions (18) need to be modified if weighted regressions are run instead of unweighted regressions.
[16] If $e_1^s = 0$, then r(1) will equal r(2).

using the period s hedonic regression as the quality adjuster. Thus the price relative defined by (24) compares the price of model 2 in period t to the price of model 1 in period s in constant utility quantity units. Hence, the period s hedonic regression may be used to express model prices in homogeneous quality adjusted units.[17]

Obviously, if the statistical agency has the period t hedonic regression available to it, then the above analysis can be repeated, with some modifications. In this case, equation (19) for period t and model 2 is:

(25) $p_2^t = f^{-1}[h^t(z_2^t,b^t) + e_2^t]$.

Recall that model 1 has the vector of characteristics z_1^s. Hence, using the period t hedonic regression, a comparable price for model 1 in period t is $f^{-1}[h^t(z_1^s,b^t)]$, the predicted period t price using the period t hedonic regression for a model with the vector of characteristics z_1^s. Thus our third estimator for an adjusted price relative for models 1 and 2 going from period s to t is:

(26) $r(3) \equiv f^{-1}[h^t(z_1^s,b^t)]/p_1^s$.

However, again, there is a problem with the use of (26) as an adjusted price relative. As above, the problem becomes apparent if $z_2^t = z_1^s$, so that the two models are in fact identical. In this case, we want our price relative to equal the actual price ratio:

(27) $p_2^t/p_1^s = f^{-1}[h^t(z_2^t,b^t) + e_2^t]/p_1^s$ using (25)
$\neq f^{-1}[h^t(z_2^t,b^t)]/p_1^s$ if $e_2^t \neq 0$.

Hence if the regression residual for model 2 in period t, e_2^t, is not equal to zero, then r(3) defined by (26) will not be an appropriate adjusted price relative. In order to compare like with like, we must multiply r(3) by an adjustment factor equal to

(28) $p_2^t/f^{-1}[h^t(z_2^t,b^t)] = f^{-1}[h^t(z_2^t,b^t) + e_2^t]/f^{-1}[h^t(z_2^t,b^t)]$.

Thus our fourth estimator r(4) for an adjusted price relative is r(3) defined by (26) times the adjustment factor defined by (28), which adjusts the period t observed price for model 2, p_2^t, onto the period t hedonic regression surface:[18]

(29) $r(4) \equiv \{f^{-1}[h^t(z_1^s,b^t)]/p_1^s\}\{p_2^t/f^{-1}[h^t(z_2^t,b^t)]\}$
$= \{p_2^t/f^{-1}[h^t(z_2^t,b^t)]\}/\{p_1^s/f^{-1}[h^t(z_1^s,b^t)]\}$.

The second expression for r(4) in (29) is again instructive. We can interpret $p_2^t/f^{-1}[h^t(z_2^t,b^t)]$ as the period t price for model 2 expressed in constant quality utility units, using the period t hedonic regression as the quality adjustment mechanism. Similarly, we can interpret $p_1^s/f^{-1}[h^t(z_1^s,b^t)]$ as the period s price for model 1 expressed in constant quality utility units, using the period t hedonic regression as the quality adjuster. Thus the price relative defined by (29) compares the price of model 2 in period t to the price of model 1 in period s in

[17] This basic idea can be traced back to Court (1939; 108) as his hedonic suggestion number one. The idea was explicitly laid out in Griliches (1971a; 59-60) (1971b; 6) and Dhrymes (1971; 111-112). It was implemented in a statistical agency sampling context by Triplett and McDonald (1977; 144).
[18] Of course, if $e_2^t = 0$, then r(3) will equal r(4).

constant utility quantity units, using the period t hedonic regression to do the quality adjustment.

If the period s and t hedonic regressions are both available to the statistical agency, then it is best to make use of *both* of the adjusted price relatives r(2) and r(4) and generate a final adjusted price relative that is a symmetric average of the two estimates.[19] Thus define our final preferred adjusted price relative r(5) as the geometric mean of r(2) and r(4):

$$(30)\ r(5) \equiv [r(2)r(4)]^{1/2}.$$

We chose the geometric mean in (30) over other simple symmetric means like the arithmetic average because the use of the geometric average leads to an adjusted price relative that will satisfy the time reversal test.[20]

Finally, suppose that period s and t hedonic regressions are not available to the statistical agency but a base period hedonic regression is available. In this case, the obvious adjusted replacement price ratio is:

$$(31)\ r(6) \equiv \{p_2^t/f^{-1}[h^0(z_2^t,b^0)]\}/\{p_1^s/f^{-1}[h^0(z_1^s,b^0)]\}.$$

Thus the price relative defined by (31) compares the price of model 2 in period t to the price of model 1 in period s in constant utility quantity units, using the period 0 hedonic regression to do the quality adjustment.

Obviously, the adjusted price relative r(5) would generally be preferable to the price relative defined by r(6), since the period 0 hedonic regression may be quite out of date if period 0 is distant from periods s and t.[21] Similar considerations suggest that more reliable results will be obtained if the chain principle is used in forming the adjusted price relatives defined by (5); i.e., the gap between the equally valid r(2) and r(4) is likely to be minimized if period s is chosen to be period t–1.[22]

In the following section, we shall assume that the statistical agency has estimated single period hedonic regressions as in this section but in addition, we assume that information on quantities sold of each model is available. Hence, Paasche, Laspeyres and superlative indexes of the type advocated by Silver and Heravi (2001) (2002a) (2002b) and Pakes (2001) can be calculated.

---

[19] Griliches (1971a; 59) noted the existence of these two equally valid estimates. Griliches (1971b; 7) also suggested taking an average of the two estimates and, as an alternative method of averaging or smoothing, he suggested using adjacent year regressions, which will be studied in sections 7 and 8 below.

[20] See Diewert (1997; 138) for an argument along these lines.

[21] Tastes will probably change over time and the characteristics domain of definition for models that exist in period 0 may be quite different from the domains of definition for the models that exist in periods s and t; i.e., the z region spanned by the period 0 hedonic regression may be quite out of date for the later periods.

[22] Our advocacy of the chain principle and of averaging equally valid results seems to be consistent with the position advocated by Griliches (1971b; 6-7): "This approach calls for relatively recent and often changing 'price' weights. Since such statistics come to us in discrete intervals, we are also faced with the usual Laspeyres-Paasche problem. The oftener we can change such weights [i.e., run a new hedonic regression], the less of a problem it will be. In practice, while one may want to use the most recent cross section to derive the relevant price weights, such estimates may fluctuate too much for comfort as the result of multicollinearity and sampling fluctuations. They should be smoothed in some way, either by choosing $w_i = (1/2)[w_i(t) + w_i(t+1)]$, or by using 'adjacent year' regressions in estimating these weights."

5. Single Period Hedonic Regressions in the Scanner Data Context

In this section, we assume that the statistical agency has both price and quantity (or value) data for the subset of the K models that are available in each period. As in the previous period, we will assume that the statistical agency has run single period hedonic regressions for periods s and t.[23]

The hedonic regression of period s can be used in order to calculate the following *Paasche type index going from period s to t*:[24]

(32) $P_P(s,t) \equiv \sum_{k \in S(t)} p_k^t q_k^t / \{\sum_{k \in [S(t) \cap S(s)]} p_k^s q_k^t + \sum_{k \in [S(t) \sim S(s)]} f^{-1}[h^s(z_k^t, b^s)] q_k^t\}$.

The summation in the numerator of the right hand side of (32) is simply the sum of price p_k^t times quantity q_k^t over all of the models k sold during period t, which is the set of indexes k represented by S(t). The first summation in the denominator of the right hand side of (32) is the product of the period s model k price, p_k^s, over all models that are present in both periods s and t while the second set of terms uses the period s estimated hedonic price of a model k that is sold in period t (which has characteristics defined by the vector z_k^t) but is not sold in period s, $f^{-1}[h^s(z_k^t, b^s)]$, times the period t quantity sold for this model, q_k^t. If we make the strong assumptions on demander's period s preferences[25] that are listed in Diewert (2001), then we can interpret $f^{-1}[h^s(z_k^t, b^s)]$ as an approximate Hicksian (1940; 114) reservation price for model k that is sold in period t but not in period s; i.e., if price is above this limiting price, then purchasers will not want to buy any units of it in period s. Thus under appropriate assumptions on consumer's preferences, the Paasche index defined by (32) will be an approximate lower bound to a theoretical Paasche-Konüs cost of living index; see Diewert (1993; 80).[26] Thus the estimated period s hedonic regression enables us to calculate a matched model type Paasche index between periods s and t, where the prices for the models that were sold in period t but not period s are filled in using the period s hedonic regression.

In a similar manner, we can use the hedonic regression for period t to fill in the missing reservation prices for models that were sold in period s but not t and we can calculate the following *Laspeyres type index going from period s to t*:[27]

(33) $P_L(s,t) \equiv \{\sum_{k \in [S(s) \cap S(t)]} p_k^t q_k^s + \sum_{k \in [S(s) \sim S(t)]} f^{-1}[h^t(z_k^s, b^t)] q_k^s\} / \{\sum_{k \in S(s)} p_k^s q_k^s\}$.

---

[23] With the availability of quantity information on the models sold, value weighted hedonic regressions of the type recommended in section 4 can be run for each period.

[24] This is Pakes' (2001; 22) Paasche complete hedonic hybrid price index. Except for error terms, it is also equal to one of Silver and Heravi's (2001) Paasche type lower bounding indexes for a true cost of living index.

[25] A stronger but simpler set of assumptions than those of Diewert (2001) are that all period s demanders of the hedonic commodity evaluate the utility of a model with characteristics vector z according to the magnitude $g^s(z)$, where $g^s(z)$ is a separable (cardinal) utility function. Under these assumptions, the equilibrium price of a model with characteristics vector z should have the period s hedonic price function equal to $g^s(z)$ times a constant. If $f^{-1}[h^s(z,\beta^s)]$ can approximate this true period s hedonic price function and if the fit of the period s hedonic regression is good so that b^s is close to β^s, then $f^{-1}[h^s(z_k^t, b^s)]$ will be an approximate Hicksian reservation price for model k that is sold in period t but not in period s.

[26] See Diewert (1993; 103-104) for an exposition of the use of Hicksian reservation prices for new and disappearing commodities in the context of Paasche and Laspeyres indexes.

[27] Except for error terms, it is equal to one of Silver and Heravi's (2001) Laspeyres type upper bounding indexes for a true cost of living index.

The summation in the denominator of the right hand side of (33) is simply the sum of price p_k^s times quantity q_k^s over all of the models k sold during period s, which is the set of indexes k represented by S(s). The first summation in the numerator of the right hand side of (33) is the product of the period t model k price, p_k^t, over all models that are present in both periods s and t while the second set of terms uses the period t estimated hedonic price of a model k that is sold in period s (which has characteristics defined by the vector z_k^s) but is not sold in period t, $f^{-1}[h^t(z_k^s,b^t)]$, times the period s quantity sold for this model, q_k^s. Under appropriate assumptions on consumer's preferences, the Laspeyres index defined by (33) will be an approximate upper bound to a theoretical Laspeyres-Konüs cost of living index; see Diewert (1993; 80). Thus the estimated period t hedonic regression enables us to calculate a matched model type Laspeyres index between periods s and t, where the prices for the models that were sold in period s but not period t are filled in using the period t hedonic regression.

If both period s and t hedonic regressions are available to the statistical agency, then since the Paasche and Laspeyres measures of price change between periods s and t are equally valid, it is appropriate to take a symmetric average of these two estimators of price change as a "final" estimator of price change between the periods.[28] As usual, we chose the geometric mean of P_L and P_P over other simple symmetric means like the arithmetic average because the use of the geometric average leads to an index that will satisfy the time reversal test.[29] Hence, define the *Fisher* (1922) *index between periods s and t* as:

$$(34)\quad P_F(s,t) \equiv [P_L(s,t)\, P_P(s,t)]^{1/2}$$

where P_P and P_L are defined by (32) and (33).[30]

It is of some interest to compute P_P, P_L and P_F defined by (32)-(34) above for the case where there are only two models: model 1, which is available in period s but not period t, and model 2, which is available in period t but not period s; i.e., we are revisiting the sampling model that was studied in section 4 above. Under these conditions, P_P defined by (32) simplifies to the following expression:

$$(35)\quad P_P(s,t) \equiv p_2^t q_2^t / f^{-1}[h^s(z_2^t,b^s)]q_2^t = p_2^t / f^{-1}[h^s(z_2^t,b^s)] = r(1)$$

where r(1) was defined in section 4 by (21). Similarly, P_L defined by (33) simplifies to the following expression:

$$(36)\quad P_L(s,t) \equiv f^{-1}[h^t(z_1^s,b^t)]q_1^s / p_1^s q_1^s = f^{-1}[h^t(z_1^s,b^t)]/p_1^s = r(3)$$

[28] If all models are present in both periods, then the Laspeyres type index defined by (33) reduces to an ordinary Laspeyres index between periods s and t and the Paasche type index defined by (32) reduces to an ordinary Paasche index. It can be seen that the weights for each of these indexes is not representative of *both* periods and hence each of the indexes (32) and (33) will be subject to substitution or representativity bias; see Diewert (2002a; 45) on the concept of representativity bias. Hence, to eliminate this bias, it is necessary to take an average of the two indexes defined by (32) and (33).

[29] See Diewert (1997; 138).

[30] An argument due originally to Konüs (1924) can be used to prove that a theoretical cost of living index lies between the Paasche and Laspeyres indexes; see also Diewert (1993; 81). However, this argument will only go through for the case where all of the characteristics are of the continuous type.

where r(3) was defined in section 4 by (26). Recall that our preferred replacement price ratios obtained in section 4 were r(2) and r(4) rather than r(1) and r(3). Hence the results obtained in this section seem to be slightly inconsistent with the results obtained in section 4.[31]

This slight inconsistency can be resolved if we make strong assumptions about the preferences of purchasers of the hedonic commodities. Suppose *all* purchasers of the hedonic commodity evaluate the relative utility of each model in period s according to the cardinal utility function $g^s(z)$ so that the relative value to purchasers of a model with characteristics vector z_1 versus a model with characteristics vector z_2 is $g^s(z_1)/g^s(z_2)$. Then in equilibrium, the period s relative price of the two models should also be $g^s(z_1)/g^s(z_2)$. Thus the period s price of a model with characteristics vector z should be proportional to $g^s(z)$. Finally, suppose that the period s econometrically estimated hedonic function, $f^{-1}[h^s(z,b^s)]$, can provide an adequate approximation to the theoretical hedonic function, $\rho^s g^s(z)$, where ρ^s is a positive constant. Under these strong assumptions, the total market utility for period s that is provided by purchases of the hedonic commodities is equal to:

$$(37)\quad Q^s \equiv \sum_{k\in S(s)} \rho^s g^s(z_k{}^s) q_k{}^s \approx \sum_{k\in S(s)} f^{-1}[h^s(z_k{}^s,b^s)] q_k{}^s$$

where we have approximated the utility to purchasers of model k in period s, $\rho^s g^s(z_k{}^s)$, by the period s hedonic regression estimated value, $f^{-1}[h^s(z_k{}^s,b^s)]$. Thus Q^s can be interpreted as the aggregate quantity of all of the models purchased in period s, where each model has been quality adjusted into constant utility units using the period s hedonic aggregator function, $g^s(z)$. In what follows, we will neglect the approximation error between lines 1 and 2 of (37) so that we identify the period s aggregate quantity purchased of the hedonic commodity, $Q^s(s)$, using the period s hedonic regression to do the quality adjustment, as follows:

$$(38)\quad Q^s(s) \equiv \sum_{k\in S(s)} f^{-1}[h^s(z_k{}^s,b^s)] q_k{}^s.$$

For each period t, we can define the *value of all models purchased* as:

$$(39)\quad V^t \equiv \sum_{k\in S(t)} p_k{}^t q_k{}^t ; \qquad t = 0,1,...,T.$$

For later reference, we also define the *period t expenditure share of model k* as follows:

$$(40)\quad s_k{}^t \equiv p_k{}^t q_k{}^t / \sum_{i\in S(t)} p_i{}^t q_i{}^t ; \qquad t = 0,1,...,T;\ k\in S(t).$$

Corresponding to the period s quantity aggregate defined by (38), we can define an *aggregate period s price level*, $P^s(s)$, by dividing $Q^s(s)$ into the period s value aggregate, V^s:

$$\begin{aligned}(41)\quad P^s(s) &\equiv V^s/Q^s(s)\\ &= V^s/\sum_{k\in S(s)} f^{-1}[h^s(z_k{}^s,b^s)] q_k{}^s && \text{using (38)}\\ &= 1/[\sum_{k\in S(s)} \{f^{-1}[h^s(z_k{}^s,b^s)]/p_k{}^s\} p_k{}^s q_k{}^s/V^s]\\ &= 1/[\sum_{k\in S(s)} \{f^{-1}[h^s(z_k{}^s,b^s)]/p_k{}^s\} s_k{}^s] && \text{using (40) for t = s}\\ &= [\sum_{k\in S(s)} s_k{}^s \{p_k{}^s/f^{-1}[h^s(z_k{}^s,b^s)]\}^{-1}]^{-1}.\end{aligned}$$

[31] We say slightly inconsistent because usually the hedonic regression observed errors $e_1{}^s$ and $e_2{}^t$ will be small and hence the differences between r(1) and r(2) and r(3) and r(4) will also be small.

Thus the aggregate period s price level using the period s hedonic regression, $P^s(s)$, is equal to a period s share weighted harmonic mean of the period s actual model prices, p_k^s, relative to the corresponding predicted period s model prices using the period s hedonic regression, $f^{-1}[h^s(z_k^s,b^s)]$.[32] Since $p_k^s = f^{-1}[h^s(z_k^s,b^s)+e_k^s]$ where e_k^s is the regression residual for model k in period s[33] and these residuals are typically close to 0 and randomly distributed around 0, it can be seen that under normal conditions, $P^s(s)$ defined by (41) will be close to 1.

Now let us use the period s hedonic regression to form a constant utility quantity aggregate for the models sold in period t. Thus model k in period t, using the estimated hedonic valuation function of period s, will have the constant utility value $f^{-1}[h^s(z_k^t,b^s)]$. Hence, the period t aggregate quantity purchased of the hedonic commodity, $Q^t(s)$, using the period s hedonic regression to do the quality adjustment into constant utility units, can be defined as follows:

$$(42)\quad Q^t(s) \equiv \sum_{k\in S(t)} f^{-1}[h^s(z_k^t,b^s)]q_k^t.$$

Corresponding to the period t quantity aggregate defined by (42), we can define an *aggregate period t price level* using the preferences of period s to do the quality adjustment, $P^t(s)$, by dividing $Q^t(s)$ into the period t value aggregate, V^t:

$$\begin{aligned}(43)\quad P^t(s) &\equiv V^t/Q^t(s)\\ &= V^t/\sum_{k\in S(t)} f^{-1}[h^s(z_k^t,b^s)]q_k^t && \text{using (42)}\\ &= 1/[\sum_{k\in S(t)} \{f^{-1}[h^s(z_k^t,b^s)]/p_k^t\}p_k^t q_k^t/V^t]\\ &= 1/[\sum_{k\in S(t)} \{f^{-1}[h^s(z_k^t,b^s)]/p_k^t\}s_k^t] && \text{using definitions (40)}\\ &= [\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^s(z_k^t,b^s)]\}^{-1}]^{-1}.\end{aligned}$$

Thus the aggregate period t price level using the period s hedonic regression, $P^t(s)$, is equal to a period t share weighted harmonic mean of the period t actual model prices, p_k^t, relative to the corresponding predicted period s model prices using the period s hedonic regression, $f^{-1}[h^s(z_k^t,b^s)]$.[34]

Having defined the period s price level $P^s(s)$ by (41) and the period t price level $P^t(s)$ by (43) using the hedonic regression of period s to do the constant utility quality adjustment, we can take the ratio of these two price levels to form a *Paasche type price index going from period s to t*, using the hedonic regression of period s, as follows:

$$\begin{aligned}(44)\quad P^{st}(s) &\equiv P^t(s)/P^s(s)\\ &= [\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^s(z_k^t,b^s)]\}^{-1}]^{-1}/[\sum_{k\in S(s)} s_k^s\{p_k^s/f^{-1}[h^s(z_k^s,b^s)]\}^{-1}]^{-1}.\end{aligned}$$

The above Paasche type index can be compared with our earlier Paasche type index defined by (32):

---

[32] It can be seen that the expression on the right hand side of (41) is a type of Paasche price index, where the price and quantity data of period s, p_k^s and q_k^s for $k\in S(s)$, act as the comparison period data and the hedonic regression period s predicted prices, $f^{-1}[h^s(z_k^s,b^s)]$ for $k\in S(s)$, act as base period prices.

[33] Our algebra here assumes that unweighted hedonic regressions have been run. If a value weighted hedonic regression has been run for period s, then the equation $p_k^s = f^{-1}[h^s(z_k^s,b^s)+e_k^s]$ must be replaced by $p_k^s = f^{-1}[h^s(z_k^s,b^s)+(v_k^s)^{-(1/2)}e_k^s]$ where the e_k^s are the residuals for the transformed period s hedonic regression.

[34] It can be seen that the expression on the right hand side of (43) is a Paasche price index, where the price and quantity data of period t, p_k^t and q_k^t for $k\in S(t)$, act as the comparison period data and the hedonic regression period s predicted prices, $f^{-1}[h^s(z_k^t,b^s)]$ for $k\in S(t)$, act as base period prices.

$$
\begin{aligned}
(45)\ P_P(s,t) &\equiv \sum_{k\in S(t)} p_k^t q_k^t / \{\sum_{k\in[S(t)\cap S(s)]} p_k^s q_k^t + \sum_{k\in[S(t)\sim S(s)]} f^{-1}[h^s(z_k^t,b^s)]q_k^t\} \\
&= V^t / \{\sum_{k\in[S(t)\cap S(s)]} p_k^s q_k^t + \sum_{k\in[S(t)\sim S(s)]} f^{-1}[h^s(z_k^t,b^s)]q_k^t\} \qquad \text{using definition (39)} \\
&= 1/\{\sum_{k\in[S(t)\cap S(s)]} [p_k^s/p_k^t]p_k^t q_k^t + \sum_{k\in[S(t)\sim S(s)]} (f^{-1}[h^s(z_k^t,b^s)]/p_k^t)p_k^t q_k^t\}/V^t \\
&= 1/\{\sum_{k\in[S(t)\cap S(s)]} [p_k^s/p_k^t]s_k^t + \sum_{k\in[S(t)\sim S(s)]} (f^{-1}[h^s(z_k^t,b^s)]/p_k^t)s_k^t\} \qquad \text{using (40)} \\
&= \{\sum_{k\in[S(t)\cap S(s)]} s_k^t [p_k^t/p_k^s]^{-1} + \sum_{k\in[S(t)\sim S(s)]} s_k^t [p_k^t/f^{-1}[h^s(z_k^t,b^s)]]^{-1}\}^{-1} \\
&= \{\sum_{k\in[S(t)\cap S(s)]} s_k^t (p_k^t/f^{-1}[h^s(z_k^t,b^s)+e_k^s])^{-1} + \sum_{k\in[S(t)\sim S(s)]} s_k^t [p_k^t/f^{-1}[h^s(z_k^t,b^s)]]^{-1}\}^{-1} \\
&\qquad \text{since } p_k^s = f^{-1}[h^s(z_k^t,b^s)+e_k^s] \text{ for } k\in S(t)\cap S(s) \\
&\approx \{\sum_{k\in[S(t)\cap S(s)]} s_k^t (p_k^t/f^{-1}[h^s(z_k^t,b^s)])^{-1} + \sum_{k\in[S(t)\sim S(s)]} s_k^t (p_k^t/f^{-1}[h^s(z_k^t,b^s)])^{-1}\}^{-1} \\
&\qquad \text{neglecting the regression residuals } e_k^s \text{ for } k\in S(t)\cap S(s) \\
&= \{\sum_{k\in S(t)} s_k^t (p_k^t/f^{-1}[h^s(z_k^t,b^s)])^{-1}\}^{-1} \\
&= P^t(s) \qquad \text{using (43).}
\end{aligned}
$$

Thus our old Paasche type index $P_P(s,t)$ is approximately equal to the numerator of our new Paasche type index $P^{st}(s)$. However, as we mentioned before, the denominator of $P^{st}(s)$, $P^s(s)$, will be approximately equal to 1, and hence, our new Paasche type index will be approximately equal to our old Paasche type index; i.e., we have

$$(46)\ P^{st}(s) \approx P_P(s,t).$$

Now consider our new Paasche type index for the case where there are only two models: model 1, which is available in period s but not period t, and model 2, which is available in period t but not period s so that we are revisiting the sampling model that was studied in section 4 above. Under these conditions, $P^{st}(s)$ defined by (44) simplifies to r(2) defined in section 4 by (24). Hence our new Paasche type index is perfectly consistent with the hedonically adjusted sampling price ratio r(2) defined earlier in section 4.

Obviously, the above analysis can be repeated except that the hedonic regression for period t is used to do the quality adjustment rather than the period s hedonic regression. Thus, we now suppose that *all* purchasers of the hedonic commodity evaluate the relative utility of each model in period t according to the cardinal utility function $g^t(z)$. Then in equilibrium, the period t price of a model with characteristics vector z should be proportional to $g^t(z)$. Suppose that the period t econometrically estimated hedonic function, $f^{-1}[h^t(z,b^t)]$, can provide an adequate approximation to the period t theoretical hedonic function, $\rho^t g^t(z)$, where ρ^t is a positive constant. Under these strong assumptions, the total market utility for period t that is provided by purchases of the hedonic commodities is equal to:

$$
\begin{aligned}
(47)\ Q^t &\equiv \sum_{k\in S(t)} \rho^t g^t(z_k^t) q_k^t \\
&\approx \sum_{k\in S(t)} f^{-1}[h^t(z_k^t,b^t)]q_k^t
\end{aligned}
$$

where we have approximated the utility to purchasers of model k in period t, $\rho^t g^t(z_k^t)$, by the period t hedonic regression estimated value, $f^{-1}[h^t(z_k^t,b^t)]$. Thus Q^t can be interpreted as the aggregate quantity of all of the models purchased in period t, where each model has been quality adjusted into constant utility units using the period t hedonic aggregator function, $g^t(z)$. In what follows, we will again neglect the approximation error between lines 1 and 2 of (47) so that we identify the period t aggregate quantity purchased of the hedonic commodity, $Q^t(t)$, using the period t hedonic regression to do the quality adjustment, as follows:

$$(48)\ Q^t(t) \equiv \sum_{k\in S(t)} f^{-1}[h^t(z_k^t,b^t)]q_k^t.$$

Corresponding to the period t quantity aggregate defined by (48), we can define an *aggregate period t price level*, $P^t(t)$, by dividing $Q^t(t)$ into the period t value aggregate, V^t:

$$
\begin{aligned}
(49)\ P^t(t) &\equiv V^t/Q^t(t) \\
&= V^t/\sum_{k\in S(t)} f^{-1}[h^t(z_k^t,b^t)]q_k^t && \text{using (48)} \\
&= 1/[\sum_{k\in S(t)} \{f^{-1}[h^t(z_k^t,b^t)]/p_k^t\}p_k^t q_k^t/V^t] \\
&= 1/[\sum_{k\in S(t)} \{f^{-1}[h^t(z_k^t,b^t)]/p_k^t\}s_k^t] && \text{using definitions (40)} \\
&= [\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^t(z_k^t,b^t)]\}^{-1}]^{-1}.
\end{aligned}
$$

Thus the aggregate period t price level using the period t hedonic regression, $P^t(t)$, is equal to a period t share weighted harmonic mean of the period t actual model prices, p_k^t, relative to the corresponding predicted period t model prices using the period t hedonic regression, $f^{-1}[h^t(z_k^t,b^t)]$.[35] Since $p_k^t = f^{-1}[h^t(z_k^t,b^t)+e_k^t]$ where e_k^t is the regression residual for model k in period t and these residuals are typically close to 0 and randomly distributed around 0, it can be seen that under normal conditions, $P^t(t)$ defined by (49) will be close to 1.

Now use the period t hedonic regression to form a constant utility quantity aggregate for the models sold in period s. Thus model k in period s, using the estimated hedonic valuation function of period t, will have the constant utility value $f^{-1}[h^t(z_k^s,b^t)]$. Hence, the *period s aggregate quantity* purchased of the hedonic commodity, $Q^s(t)$, using the period t hedonic regression to do the quality adjustment into constant utility units, can be defined as follows:

$$(50)\ Q^s(t) \equiv \sum_{k\in S(s)} f^{-1}[h^t(z_k^s,b^t)]q_k^s.$$

Corresponding to the period s quantity aggregate defined by (50), we can define an *aggregate period s price level* using the preferences of period t to do the quality adjustment, $P^s(t)$, by dividing $Q^s(t)$ into the period s value aggregate, V^s:

$$
\begin{aligned}
(51)\ P^s(t) &\equiv V^s/Q^s(t) \\
&= V^s/\sum_{k\in S(s)} f^{-1}[h^t(z_k^s,b^t)]q_k^s && \text{using (50)} \\
&= 1/[\sum_{k\in S(s)} \{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}p_k^s q_k^s/V^s] \\
&= 1/[\sum_{k\in S(s)} \{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}s_k^s] && \text{using definitions (40)} \\
&= [\sum_{k\in S(s)} s_k^s\{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}]^{-1}.
\end{aligned}
$$

Thus the aggregate period s price level using the period t hedonic regression, $P^s(t)$, is equal to the reciprocal of a period s share weighted arithmetic mean of the predicted period s model prices in period t using the period t hedonic regression, $f^{-1}[h^t(z_k^s,b^t)]$, relative to the period s actual model prices, p_k^s.[36]

Having defined the period s price level $P^s(t)$ by (51) and the corresponding period t price level $P^t(t)$ by (49) using the hedonic regression of period t to do the constant utility quality

[35] It can be seen that the expression on the right hand side of (49) is a type of Paasche price index, where the price and quantity data of period t, p_k^t and q_k^t for $k\in S(t)$, act as the comparison period data and the hedonic regression period t predicted prices, $f^{-1}[h^t(z_k^t,b^t)]$ for $k\in S(t)$, act as base period prices.

[36] It can be seen that the expression on the right hand side of (51) is the reciprocal of a kind of Laspeyres price index, where the price and quantity data of period s, p_k^s and q_k^s for $k\in S(s)$, act as the base period price and quantity data and the hedonic regression period t predicted prices, $f^{-1}[h^t(z_k^s,b^t)]$ for $k\in S(s)$, act as comparison period prices.

adjustment, we can take the ratio of these two price levels to form a *Laspeyres type price index going from period s to t*, using the hedonic regression of period t, as follows:

$$
\begin{aligned}
(52)\ P^{st}(t) &\equiv P^t(t)/P^s(t) \\
&= [\textstyle\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^t(z_k^t,b^t)]\}^{-1}]^{-1}/[\sum_{k\in S(s)} s_k^s\{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}]^{-1} \\
&= [\textstyle\sum_{k\in S(s)} s_k^s\{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}][\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^t(z_k^t,b^t)]\}^{-1}]^{-1} \\
&\approx \textstyle\sum_{k\in S(s)} s_k^s\{f^{-1}[h^t(z_k^s,b^t)]/p_k^s\}
\end{aligned}
$$

where the last line above follows from the assumption that $[\sum_{k\in S(t)} s_k^t\{p_k^t/f^{-1}[h^t(z_k^t,b^t)]\}^{-1}]^{-1}$ will be approximately equal to 1.[37]

The above Laspeyres type index can be compared with our earlier Laspeyres type index defined by (33):

$$
\begin{aligned}
(53)\ P_L(s,t) &\equiv \{\textstyle\sum_{k\in[S(s)\cap S(t)]} p_k^t q_k^s + \sum_{k\in[S(s)\sim S(t)]} f^{-1}[h^t(z_k^s,b^t)]q_k^s\}/\{\sum_{k\in S(s)} p_k^s q_k^s\} \\
&= \{\textstyle\sum_{k\in[S(s)\cap S(t)]} p_k^t q_k^s + \sum_{k\in[S(s)\sim S(t)]} f^{-1}[h^t(z_k^s,b^t)]q_k^s\}/V^s \qquad \text{using (39) for t = s} \\
&= \{\textstyle\sum_{k\in[S(s)\cap S(t)]} [p_k^t/p_k^s]p_k^s q_k^s + \sum_{k\in[S(s)\sim S(t)]} (f^{-1}[h^t(z_k^s,b^t)]/p_k^s)p_k^s q_k^s\}/V^s \\
&= \textstyle\sum_{k\in[S(s)\cap S(t)]} [p_k^t/p_k^s]s_k^s + \sum_{k\in[S(s)\sim S(t)]} (f^{-1}[h^t(z_k^s,b^t)]/p_k^s)s_k^s \qquad \text{using (40)} \\
&= \textstyle\sum_{k\in[S(s)\cap S(t)]} [f^{-1}[h^t(z_k^t,b^t)+e_k^t]/p_k^s]s_k^s + \sum_{k\in[S(s)\sim S(t)]} (f^{-1}[h^t(z_k^s,b^t)]/p_k^s)s_k^s \\
&\qquad \text{since } p_k^t = f^{-1}[h^t(z_k^t,b^t)+e_k^t] \text{ for } k\in S(s)\cap S(t) \\
&\approx \textstyle\sum_{k\in[S(s)\cap S(t)]} [f^{-1}[h^t(z_k^t,b^t)]/p_k^s]s_k^s + \sum_{k\in[S(s)\sim S(t)]} (f^{-1}[h^t(z_k^s,b^t)]/p_k^s)s_k^s \\
&\qquad \text{neglecting the regression residuals } e_k^t \text{ for } k\in S(s)\cap S(t) \\
&= 1/P^s(t) \qquad \text{using definition (51)} \\
&\approx P^{st}(t) \qquad \text{since } P^t(t) \text{ is approximately equal to 1.}
\end{aligned}
$$

Thus our old Laspeyres type index $P_L(s,t)$ is approximately equal to our new Laspeyres type index $P^{st}(t)$.

Consider our new Laspeyres type index for the case where there are only two models: model 1, which is available in period s but not period t, and model 2, which is available in period t but not period s so that we are revisiting the sampling model that was studied in section 4 above. Under these conditions, $P^{st}(t)$ defined by (52) simplifies to r(4) defined in section 4 by (29). Hence our new Laspeyres type index, $P^{st}(t)$, is perfectly consistent with the hedonically adjusted sampling price ratio r(4) defined earlier in section 4.

As usual, if hedonic regressions are available for both periods s and t, then the two indexes $P^{st}(s)$ and $P^{st}(t)$, defined by (44) and (52) respectively, should be averaged geometrically to form a final Fisher type estimate of price change going from period s to t.

We now turn our attention to bilateral hedonic regressions (i.e., hedonic regressions that involve the data of two periods instead of only one period) that also make use of a time dummy variable.

[37] The last line on the right hand side of (52) is the hedonic index that is advocated by Pakes (2001; 26). Pakes assumes that s = t–1.

6. Unweighted Bilateral Hedonic Regressions with Time as a Dummy Variable

We now consider the following hedonic regression model, which utilizes the data of periods s and t:

$$(54)\ f(p_k^s) = \beta_0 + \sum_{n=1}^{N} f_n(z_{kn}^s)\beta_n + \varepsilon_k^s ; \qquad k \in S(s);$$
$$(55)\ f(p_k^t) = \gamma_{st} + \beta_0 + \sum_{n=1}^{N} f_n(z_{kn}^t)\beta_n + \varepsilon_k^t ; \qquad k \in S(t);$$

where ε_k^s and ε_k^t are independently distributed error terms with mean 0 and variance σ^2, f(x) is either the identity function f(x) ≡ x or the natural logarithm function f(x) ≡ ln x and the functions of one variable f_n are either the identity function, the logarithm function or a dummy variable which takes on the value 1 if the characteristic n is present in model k or 0 otherwise. Note that the β regression coefficients in (54) are constrained to be the same as the corresponding β coefficients in (55). Note also that equations (55) have added a time dummy variable, γ_{st}, and this coefficient will summarize the overall price change in the various models going from period s to t.[38]

Before proceeding further, we briefly discuss some of the advantages and disadvantages of the dummy variable model defined by (54) and (55) versus running separate single period regressions of the type defined by (1) for periods s and t and then using these separate regressions to form two separate estimates of quality adjusted prices which would be averaged in some way in order to form an overall measure of price change between periods s and t. The main advantage of the latter method is that it is more flexible; i.e., changes in tastes between periods can readily be accommodated. However, this method has the disadvantage that *two* distinct estimates of period s to t price change will be generated by the method (one using the regression for period s and the other using the regression for period s) and it is somewhat arbitrary how these two estimates are to be averaged to form a single estimate of price change. The main advantages of the dummy variable method are that it conserves degrees of freedom and is less subject to multicollinearity problems[39] and there is no ambiguity about the measure of overall price change between periods s and t.[40]

We have considered only the case of two periods since this is the case of most interest to statistical agencies who must provide measures of price change between two periods.

---

[38] This two period time dummy variable hedonic regression (and its extension to many periods) was first considered explicitly by Court (1939; 109-111) as his hedonic suggestion number two. Court (1939; 110) chose to transform the prices by the log transformation on empirical grounds: "Prices were included in the form of their logarithms, since preliminary analysis indicated that this gave more nearly linear and higher simple correlations." Court (1939; 111) then used adjacent period time dummy hedonic regressions as links in a longer chain of comparisons extending from 1920 to 1939 for US automobiles: "The net regressions on time shown above are in effect price link relatives for cars of constant specifications. By joining these together, a continuous index is secured." If the two periods being compared are consecutive periods, Griliches (1971b; 7) coined the term "adjacent year regression" to describe this dummy variable hedonic regression model.

[39] This advantage was noted by Griliches (1971b; 8): "The time dummy approach does have the advantage, if the comparability problem can be solved, of allowing us to ignore the ever present problem of multicollinearity among the various dimensions."

[40] Griliches (1971b; 7) has the following very nice summary justification for the use of the time dummy variable method: "The justification for this [method] is very simple and appealing: we allow as best we can for all of the major differences in specifications by 'holding them constant' through regression techniques. That part of the average price change which is not accounted for by any of the included specifications will be reflected in the coefficient of the time dummy and represents our best estimate of the 'unexplained-by-specification-change average price change."

However, the bilateral model defined by (54) and (55) can encompass both the fixed base situation (where s will equal the base period 0) or the chained situation where s will equal t–1. It is also of interest to consider the two period case because in this situation, we can draw on many of the ideas that have been introduced into bilateral index number theory, which also deals with the problem of measuring price change between two periods.

We first consider the case where f is the identity transformation. Let us estimate the unknown parameters in (54) and (55) by least squares regression and denote the estimates for the β_n by b_n for n = 0,1,....,N and the estimate for γ_{st} by c_{st}. Denote the least squares residuals for equations (54) and (55) with f defined to be the identity transformation by e_k^s and e_k^t respectively. Then we have the following equations, which relate the model prices in the two periods to their predicted values and the sample residuals:

$$(56)\ p_k^s = b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n + e_k^s ; \qquad k \in S(s);$$
$$(57)\ p_k^t = c_{st} + b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n + e_k^t ; \qquad k \in S(t).$$

Now consider a hypothetical situation where the models sold during periods s and t are exactly the same so that there are say K common models pertaining to the two periods. Suppose further that the model prices in period t are all exactly λ times greater than the corresponding model prices in period s, where λ is a positive constant. Under these conditions, it seems reasonable to ask that the regression predicted values for the period t models be exactly equal to λ times the regression predicted values for the same models in period s; i.e., we want the following equations to be satisfied:

$$(58)\ c_{st} + b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n = \lambda[\, b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n] ; \qquad k = 1,...,K.$$

In general, if $K > N+2$ and $\lambda \neq 1$, it can be seen that equations (56) cannot be solved for any coefficients c_{st}, b_0, b_1,...,b_N. Hence, our conclusion is that the linear time dummy hedonic regression model defined by (56) and (57) is not a very good one, since it will not give us the "right" answer in a simple situation where all model prices are proportional for the two periods.[41] Of course, this homogeneity problem with the linear dummy variable regression model can be solved if we replace equations (57) by the following equations:

$$(59)\ p_k^t = c_{st}[b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n] + e_k^t ; \qquad k \in S(t).$$

In equations (59), the time dummy variable, c_{st}, now appears in a multiplicative fashion. Thus, the problem with the estimating equations (56) and (59) is that we no longer have a linear regression model; nonlinear estimation techniques would have to be used.

Since nonlinear regression models are more difficult to estimate and may suffer from reproducibility problems, we will turn our attention to the second set of bilateral hedonic regression models, where f is the log transformation. In this case, the counterparts to equations (56) and (57) are the following equations:

$$(60)\ \ln p_k^s = b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n + e_k^s ; \qquad k \in S(s);$$
$$(61)\ \ln p_k^t = c_{st} + b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n + e_k^t ; \qquad k \in S(t).$$

[41] Diewert (2001) also argued on theoretical grounds that dummy variable hedonic regression models that used untransformed prices as dependent variables did not have good properties.

Exponentiating both sides of (60) and (61) leads to the following equations that will be satisfied by the data and the least squares estimators for (60) and (61):

(62) $p_k^s = \exp[b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n]\exp[e_k^s]$ $\quad k \in S(s)$;
(63) $p_k^t = \exp[c_{st}]\exp[b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n]\exp[e_k^t]$; $\quad k \in S(t)$.

Again consider a hypothetical situation where the models sold during periods s and t are exactly the same so that there are K common models pertaining to the two periods. Again suppose that the model prices in period t are all exactly λ times greater than the corresponding model prices in period s, where λ is a positive constant. Again we ask that the regression predicted values for the period t models be exactly equal to λ times the regression predicted values for the same models in period s; i.e., we want the following equations to be satisfied:

(64) $\exp[c_{st}]\exp[b_0 + \sum_{n=1}^N f_n(z_{kn}^t)b_n] = \lambda\{\exp[b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n]\}$; $\quad k = 1,...,K.$

It can be seen that if we choose $c_{st} = \ln \lambda$, then we can satisfy equations (64). Hence we conclude (from a test approach perspective) that it is preferable to run linear bilateral dummy variable hedonic regressions using the log transformation for the dependent variable rather than leaving the model prices untransformed. Thus, we have again reinforced the case for using the log transformation on the dependent variable in hedonic regression models.

The bilateral log hedonic regression model is defined by (54) and (55) where f is the log transformation. It can be seen that in this case, the theoretical index of price change going from period s to t is $\exp[\gamma_{st}]$ and the sample estimator of this population measure is:

(65) $P(s,t) \equiv \exp[c_{st}]$

where c_{st} is the least squares estimator for the shift parameter γ_{st}. Note that we put the shift parameter in equations (55) rather than in equations (54). The choice of base period should not matter so let us consider the following bilateral log regression model which puts the shift parameter γ_{ts} in the period s equations rather than in the period t equations:

(66) $\ln p_k^s = \gamma_{ts} + \beta_0^* + \sum_{n=1}^N f_n(z_{kn}^s)\beta_n^* + \varepsilon_k^s$; $\quad k \in S(s)$;
(67) $\ln p_k^t = \beta_0^* + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n^* + \varepsilon_k^t$; $\quad k \in S(t)$.

Denote the least squares estimates for β_n^* by b_n^* for n = 0,1,....,N and the estimate for γ_{ts} by c_{ts}. For the regression model defined by (66) and (67), it can be seen that the theoretical index of price change going from period t to s is $\exp[\gamma_{ts}]$ and the sample estimator of this population measure is:

(68) $P(t,s) \equiv \exp[c_{ts}]$.

The question now is: how does P(s,t) defined by (65) relate to P(t,s) defined by (68)? Ideally, we would like these two estimators of price change to satisfy the following *time reversal test*:

(69) $P(t,s) = 1/P(s,t)$.

If we compare the original log linear regression model defined by (54) and (55) (with f being the log transformation) with the new model defined by (66) and (67), it can be seen that the

right hand side exogenous variables are identical except that γ_{ts} appears in the first set of equations in (66) and (67) while γ_{st} appears in the second set of equations in (54) and (55). The transpose of the column in the X matrix that corresponds to γ_{ts} in (66) and (67) is equal to $[1_1^T,0_2^T]$ where 1_1 is a column vector of ones of dimension equal to the number of models in the set S(s) and 0_2 is a column vector of zeros of dimension equal to the number of models in the set S(t). The transpose of the column in the X matrix that corresponds to γ_{st} in (54) and (55) is equal to $[0_1^T,1_2^T]$ where 0_1 is a column vector of zeros of dimension equal to the number of models in the set S(s) and 1_2 is a column vector of ones of dimension equal to the number of models in the set S(t). However, note that both models have the constant term β_0 (or β_0^*) in every equation and the transpose of the column in the X matrix that corresponds to this constant term is equal to $[1_1^T,1_2^T]$ in both models. It can be seen that the subspace spanned by the X columns corresponding to β_0 and γ_{st} in (54) and (55) is equal to the subspace spanned by the X columns corresponding to β_0^* and γ_{ts} in (66) and (67) and the two sets of parameters are related by the following equations:

(70) $[0_1^T,1_2^T]\,\gamma_{st} + [1_1^T,1_2^T]\,\beta_0 = [1_1^T,0_2^T]\,\gamma_{ts} + [1_1^T,1_2^T]\,\beta_0^*$.

Equations (70) are equivalent to the following 2 equations in the four variables γ_{st}, β_0, γ_{ts} and β_0^*:

(71) $0\,\gamma_{st} + 1\,\beta_0 = 1\,\gamma_{ts} + 1\,\beta_0^*$;
$1\,\gamma_{st} + 1\,\beta_0 = 0\,\gamma_{ts} + 1\,\beta_0^*$.

Thus given γ_{st} and β_0, the corresponding γ_{ts} and β_0^* can be obtained using equations (71) as:

(72) $\gamma_{ts} = -\gamma_{st}$; $\beta_0^* = \gamma_{st} + \beta_0$.

Equations (72) also hold for the least squares estimators for the two hedonic regression models. In particular, we have:

(73) $c_{ts} = -c_{st}$.

Hence, exponentiating both sides of (73) gives us $\exp[c_{ts}] = 1/\exp[c_{ts}]$ and this equation is equivalent to (69) using definitions (68) and (65). Thus we have shown that the estimator of price change P(s,t) defined by (65) (which corresponds to the least squares estimators of the initial log hedonic regression model defined by (54) and (55) with $f(p) \equiv \ln p$) is equal to the reciprocal of the estimator of price change P(t,s) defined by (68) (which corresponds to the second log hedonic regression model defined by (66) and (67) so that the two bilateral dummy variable hedonic regressions satisfy the time reversal test (69).

The results in this section strongly support the use of the logarithms of model prices as the dependent variables in an unweighted bilateral hedonic regression model with a time dummy variable. In the following section, we will study the properties of *weighted* bilateral hedonic regression models.

7. Weighted Bilateral Hedonic Regressions with Time as a Dummy Variable

Given the results in the previous section, we consider only weighted bilateral hedonic regressions that use the log of model prices as the dependent variable, before weighting the

equations. We also draw on the results in section 3 and consider only value weighting. Thus we now consider the following *value weighted hedonic regression model*, which utilizes the data of periods s and t:

$$(74)\ (v_k^s)^{1/2} \ln p_k^s = (v_k^s)^{1/2}[\beta_0 + \sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s ; \qquad k \in S(s);$$
$$(75)\ (v_k^t)^{1/2} \ln p_k^t = (v_k^t)^{1/2}[\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t ; \qquad k \in S(t);$$

where the model sales values for period t, v_k^t, were defined by (14) and ε_k^s and ε_k^t are independently distributed error terms with mean 0 and variance σ^2.

The weighted model defined by (74) and (75) is the bilateral counterpart to our single equation weighted hedonic regression model that was studied in section 3 above. However, in the present bilateral context, we now encounter a problem that was absent in the single equation context. The problem is this: if there is high inflation going from period s to t, then the period t model sales values v_k^t can be very much bigger than the corresponding period s model sales values v_k^s due to this general inflation. Hence, the assumption of homoskedastic residuals between equations (74) and (75) is unlikely to be satisfied. Hence, it is necessary to pick new weights that will eliminate this problem.

Our tentative initial suggested solution to the above problem caused by general inflation between the two periods is to use the model expenditure shares, s_k^s and s_k^t defined earlier by (40) as the weights in (74) and (75) in place of the model expenditures, v_k^s and v_k^t. Thus we recommend the use of the following *expenditure share weighted hedonic regression model*, which utilizes the data of periods s and t:[42]

$$(76)\ (s_k^s)^{1/2} \ln p_k^s = (s_k^s)^{1/2}[\beta_0 + \sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s ; \qquad k \in S(s);$$
$$(77)\ (s_k^t)^{1/2} \ln p_k^t = (s_k^t)^{1/2}[\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t ; \qquad k \in S(t);$$

where ε_k^s and ε_k^t are independently distributed error terms with mean 0 and variance σ^2.

Denote the least squares estimates for β_n by b_n for n = 0,1,....,N and the estimate for γ_{st} by c_{st}. For the regression model defined by (76) and (77), it can be seen that the theoretical index of price change going from period t to s is $\exp[\gamma_{st}]$ and the sample estimator of this population measure is:

$$(78)\ P_1(s,t) \equiv \exp[c_{st}].$$

It can be shown that $P_1(s,t)$ defined by (78) in this section has the same desirable property that P(s,t) defined by (65) in the previous section had: namely, if the models are identical in the two periods (and the model expenditure shares are identical for the two periods) and the model prices in period t are all exactly λ times greater than the corresponding model prices in period s, then $P_1(s,t)$ is exactly equal to λ.[43]

[42] Diewert (2002b) considered a model similar to (76) and (77) except that all of the explanatory variables were dummy variables and showed that weighting by the square roots of expenditure shares led to a very reasonable index number formula to measure the price change between the two periods. Thus the model defined by (76) and (77) is consistent with his results.

[43] See Proposition 1 in the Appendix.

The restriction that the expenditure shares be identical in the two periods in the identical model case is a bit unrealistic. Moreover, in the identical models case, it would be nice if $P_1(s,t)$ defined by (78) turned out to equal the Törnqvist price index, since this index is a preferred one from the viewpoints of both the stochastic and economic approaches to index number theory.[44] Hence in place of the model defined by (76) and (77), when a model is present in both periods, let us use the *average* sales share for that model, $(1/2)(s_k^s+s_k^t)$, as the weight for that model in both periods. In this revised weighting scheme, the old period s equations (76) are replaced by the following two sets of equations:

$$(79)\ (s_k^s)^{1/2} \ln p_k^s = (s_k^s)^{1/2}[\beta_0 + \sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s ; \qquad k \in [S(s) \sim S(t)];$$
$$(80)\ [(1/2)(s_k^s+s_k^t)]^{1/2} \ln p_k^s = [(1/2)(s_k^s+s_k^t)]^{1/2} [\beta_0 + \sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s ; \ k \in S(s) \cap S(t).$$

Thus if a model k is present in period s but not present in period t, then we use the square root of the period s sales share for that model, $(s_k^s)^{1/2}$, as the weight, which means this model is included in equations (79). On the other hand, if model k is present in both periods, , then we use the square root of the arithmetic average of the period s and t sales shares for that model, $[(1/2)(s_k^s+s_k^t)]^{1/2}$, as the weight, which means this model is included in equations (80). Similarly, the old period s equations (77) are replaced by the following two sets of equations:

$$(81)\ (s_k^t)^{1/2} \ln p_k^t = (s_k^t)^{1/2}[\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t ; \qquad k \in [S(t) \sim S(s)];$$
$$(82)\ [(1/2)(s_k^s+s_k^t)]^{1/2} \ln p_k^t = [(1/2)(s_k^s+s_k^t)]^{1/2} [\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t ;$$
$$k \in S(s) \cap S(t).$$

Thus if a model k is present in period t but not present in period s, then we use the square root of the period t sales share for that model, $(s_k^t)^{1/2}$, as the weight, which means this model is included in equations (81). On the other hand, if model k is present in both periods, then we use the square root of the arithmetic average of the period s and t sales shares for that model, $[(1/2)(s_k^s+s_k^t)]^{1/2}$, as the weight, which means this model is included in equations (82). As usual, we assume that ε_k^s and ε_k^t are independently distributed error terms with mean 0 and variance σ^2.

Denote the least squares estimates for β_n by b_n for n = 0,1,....,N and the estimate for γ_{st} by c_{st}. For the regression model defined by (79)-(82), it can be seen that the theoretical index of price change going from period t to s is $\exp[\gamma_{st}]$ and the sample estimator of this population measure is:

$$(83)\ P_2(s,t) \equiv \exp[c_{st}].$$

It can be shown that $P_2(s,t)$ defined by (83) has the following desirable property: if the models are identical in the two periods, then $P_2(s,t)$ is equal to the Törnqvist price index between the two periods.[45] Hence *it appears that the weighted hedonic regression model defined by (79)-(82) is a "natural" weighted hedonic regression model that provides a generalization of the Törnqvist price index to cover the case where the models are not matched.* If there are no models in common for the two periods under consideration, then the model defined by (79)-(82) reduces to our earlier model defined by (76)-(77).

[44] See Diewert (2002a).
[45] This follows from Corollary 5.2 in the Appendix.

As in the previous section, it is somewhat arbitrary whether we put the time dummy variable in the period t equations or whether we put it in the period s equations. If we put the time dummy in the period s equations as the parameter γ_{ts} and obtain a weighted least squares estimate c_{ts} for this population parameter, the theoretical index of price change going from period t to s is $\exp[\gamma_{ts}]$ and the sample estimator of this population measure is:

(84) $P^*(t,s) \equiv \exp[c_{ts}]$.

As in the previous section, we would like $P^*(t,s)$ to equal the reciprocal of P(s,t). It turns out that this property is true for the weighted hedonic regressions defined by (76) and (77) and (79)-(82) in this section as well as for the unweighted ones defined in the previous section; see Proposition 4 in the Appendix. Hence it does not matter whether we put the time dummy variable in period s or t: our measure of overall price change between the two periods will be invariant to this choice for the two weighted hedonic regressions considered in this section.

Using the results in the Appendix, we can also show that $P_1(s,t)$ and $P_2(s,t)$ both satisfy the identity test (A6), the homogeneity tests (A4) and (A5) and the time reversal test (A7) as we have already indicated. Thus both of these hedonic price indexes have some good a priori properties.

Which bilateral weighted hedonic index is best? From the viewpoint of representativity, $P_1(s,t)$ seems best: the models present in each period are weighted by expenditure shares that pertain to that period. However, the loss of representativity for $P_2(s,t)$ is probably not large in most applications and $P_2(s,t)$ has the advantage of being consistent with the use of a Törnqvist price index in the matched models case.[46] Thus at this stage of research, we lean towards the use of $P_2(s,t)$.

We turn now to a discussion of the treatment of regression outliers.

8. Outliers and Influence Analysis

In the context of traditional sampling techniques used by statistical agencies, usually provision is made for the deletion of outliers in the samples of prices collected. This raises the issue as to whether outliers should also be deleted in the hedonic regression context.

In the unweighted context, the deletion of sample outlier observations should be permitted. Since influence analysis is just an extension of outlier analysis (an influential observation is one which greatly influences the estimated regression coefficients), the deletion of influential observations should also be permitted.[47]

However, in the weighted context, the situation is somewhat different for two reasons.

Assuming that we have complete market information on the prices and quantities sold of all models being considered for the two periods under consideration, then we are in the same situation as we would be if we were applying traditional bilateral index number theory. In

[46] Moreover, in the matched models case, $P_2(s,t)$ has the advantage of being independent of the hedonic regression coefficients $b_0,b_1,...,b_N$, whereas $P_1(s,t)$ is not.

[47] In the unweighted bilateral hedonic regression context, we need only delete observations that influence the estimate of γ_{st} since the other parameters are not of great significance in this context. For an exposition of the various approaches to influence analysis, see Chapter 4 in Chatterjee and Hadi (1988).

this latter context, (assuming reliable data), traditional index number theory does not suggest dropping out prices and quantities that look a bit unusual.[48] Thus the dropping of outliers or influential observations in the context of running a hedonic regression could lead to results that would not be comparable to the results of say a maximum overlap superlative index.

Our second reason for advocating a bit of caution in dropping outliers or influential observations in the weighted hedonic regression context is that in this weighted context, an individual observation does not have equal weight! Consider the case of a hugely popular model that accounts for almost all of the sales in a given period. Dropping this weighted observation would frequently lead to a big change in the weighted regression but on representativity grounds, we would not want to drop this observation. Hence traditional outlier and influence analysis would have to be somehow adapted to deal with this situation. Until this new methodology is available, I would urge a cautious approach to the dropping of observations in the context of weighted hedonic regressions.

9. Do the Signs of Hedonic Regression Coefficients Matter?

A recent paper by Pakes (2001) has stimulated a certain amount of controversy in the literature on hedonic regressions. Hulten (2002) has provided a nice summary of the more controversial parts of Pakes' analysis and Hulten labels these controversial Propositions as Pakes I, Pakes II and Pakes III. We will follow Hulten's interpretation of Pakes below.

Hulten (2002; 23) states the *Pakes Proposition I* as follows: the hedonic function is equal to a producers' marginal cost function plus a market power function that depends on the elasticities of demand for characteristics.[49]

Hulten's (2002; 25) *Pakes Proposition II* is the following corollary to Proposition I: the price of a product can go *down* when it acquires *more* of a given characteristic. In other words, the sign of a hedonic coefficient for a characteristic can go in the "wrong" direction![50]

Hulten's (2002; 25) *Pakes Proposition III* is that the two single period hedonic regressions pertaining to any two periods being compared may not bear any close relationship to each other (this follows from the first Proposition that implies that changing market power between the two periods might lead to quite different hedonic regressions) but only one of the two regressions is required to undertake a quality adjustment.[51]

---

[48] This is not quite true since there is some literature on measuring core inflation that suggests the deletion of outliers. However, the goal of this literature is usually to obtain better estimates of trend or future inflation. Traditional index numbers that do not drop observations are still acceptable as measures of past inflation.

[49] Pakes (2001; 10) distinguishes between the competitive case and the more normal market power case as follows: "That is, in the marginal cost pricing equilibrium the hedonic function *is* the marginal cost function. However in the Bertrand equilibrium the hedonic function is the sum of the marginal cost function and a function that summarizes the relationship between markups and characteristics."

[50] "Hedonic regressions have been used in research for some time and they are often found to have coefficients which are 'unstable' either over time or across markets, and which clash with the naive intuition that the 'marginal willingness to pay for a characteristic equaled its marginal cost of production'. I hope this discussion has made it amply clear that these models can be *very misleading*. The derivatives of a hedonic price function should not be interpreted as either willingness to pay derivatives or cost derivatives; rather they are formed from a complex equilibrium process." Ariel Pakes (2001; 14).

[51] The hedonic indexes that Pakes (2001; 26) considers are all of the type defined by the last line of (52) above where the chain principle is used so that $s = t-1$.

We discuss each of these Propositions below.

I agree with Pakes that the determination of model prices in any one period is determined by the interaction of demander's preferences, the costs of model suppliers and the degree of market power of suppliers. However, I disagree with Pakes' contention that the hedonic function *must* be equal to marginal cost function plus a markup function. It seems to me that Pakes is viewing the hedonic function through the eyes of producers when what is required is a consumer view. After all, the purpose of the hedonic exercise is to find how *demanders* (and *not* suppliers) of the product value alternative models in a given period.[52] Thus for the present purpose, it is the preferences of consumers that should be decisive, and *not* the technology and market power of producers. The situation is similar to ordinary general equilibrium theory where equilibrium price and quantity for each commodity is determined by the interaction of consumer preferences and producer's technology sets and market power. However, there is a big branch of applied econometrics that ignores this complex interaction and simply uses information on the prices that consumers face, the quantities that they demand and perhaps demographic information in order to estimate systems of consumer demand functions. Then these estimated demand functions are used to form estimates of consumer utility functions and these functions are often used in applied welfare economics. What producers are doing is entirely irrelevant to these exercises in applied econometrics with the exception of the prices that they are offering to sell at. In other words, we do not need information on producer marginal costs and markups in order to estimate consumer preferences: all we need are selling prices.[53] I believe that the situation is similar in the context of estimating a hedonic price function for a given period. For welfare purposes, we need to assume that the hedonic model price function is proportional to a (separable from everything else) hedonic utility function that gives the utility that demanders will get as a function of model characteristics. We then make the heroic assumption that the actual prices of models that were sold during the period under consideration are proportional to this assumed hedonic utility function.

---

[52] This position seems consistent with the position of Griliches (1971b; 14), who argued that it is the user value or utility of a model that is the "right" characteristic for government statisticians to attempt to measure: "Most economists would agree that they would like the 'price' index to be a 'price-of-living' or 'utility' indicator. Many government statisticians in charge of producing actual price indexes will reply that the *cannot* achieve this and that therefore they should not even try, but should concentrate instead on some more 'objective' index of 'transaction' prices and/or allow only for those 'quality' changes which are based on 'production' costs. The fact that 'truth' cannot be achieved doesn't mean that one shouldn't strive to do so, though I sympathize with the position that it is better to measure well something definite than to do a very poor job on a more interesting but also more nebulous concept. Nevertheless, I would deny the contention that 'transaction' units or 'production' costs are much more definitive concepts. In general, they too make little sense without some appeal to utility considerations." Griliches (1971b; 14-15) went on to definitely reject the production cost viewpoint of adjusting for quality changes: "Nor are 'production costs' an adequate guide to quality changes without a check of their utility implications. ... Nor should we ignore 'costless' changes if we can measure them. If the consumer is in fact buying 'horsepower', and if a design change makes it possible to deliver more horsepower from the same size and 'cost' engine, then the price of horsepower to the consumer has fallen *and he is better off*."

[53] Hulten (2002; 26) also comments on the similarity of hedonic regression estimation with the estimation of conventional supply and demand functions: "Indeed, the Pakes II result has precedent in conventional price-quantity analysis. When the price of a good is regressed on its quantity, it is well known that the underlying supply and demand curves generally cannot be identified separately, and that the regression coefficients will be unstable and can easily have the 'wrong' sign." This is true as far as it goes but what enables consumer demand analysis to "succeed" in this situation is that we do not estimate a single demand function but rather estimate a system of demand functions, where an exogenous identifying variable is "income". If we were using the same price and quantity data to estimate a system of producer supply functions, there would be different exogenous variables appearing in the producer system such as capital and labor used by the producers.

It is clear that there are a number of problems with the above assumptions:

- The separability assumption is very unrealistic.

- Different demanders may have very different hedonic utility functions and so over time as characteristic costs and markups change, different classes of consumers may be induced to enter the market and thus the estimated hedonic utility function may be quite unstable over time.

- The assumption that the market is in equilibrium is also suspect, particularly in the context of new products, where it takes time for demanders to discover the new products.

- Another consequence of the equilibrium assumptions that we have made is that all models which are sold in a given period are equally desirable; i.e., they all yield equal utility per dollar spent. But in practice, some models are vastly more popular than others and our suggested approach does not directly take this fact into account.[54]

In spite of the above problems, I believe that the consumer valuation approach is more appropriate in the context of making quality adjustments for CPI purposes than the producer valuation approach proposed by Pakes. Note that there do not appear to be any welfare implications whatsoever in making hedonic price adjustments using the framework of Pakes.

We turn now to Pakes Proposition II; namely, that the price of a product can go *down* when it acquires *more* of a given characteristic. In other words, the signs of hedonic regression coefficients do not have to be sign restricted.[55] In evaluating this Proposition, it is again necessary to keep in mind the purpose for running the hedonic regression, which is to provide utility valuations for possibly hypothetical models that are sold in one period but not in the other period. For this welfare economics type purpose, I believe that it makes sense to impose a priori sign restrictions on the regression coefficients.[56] Hence if we believe that demanders of a model will, on average, get a higher utility from a model that has more of an a priori desirable feature, then we should make sure that our estimated hedonic regression does not contradict these a priori beliefs.[57] Thus we are taking the point of view here that we impose our theory on the data and do the econometric estimation which fits the data best, consistent with our prior beliefs. An alternative (but perfectly defensible point of view) is that we use the data to discover whether our a priori theories are consistent with the data.[58] However, it

[54] However the use of expenditure or share weighted hedonic regressions can partially overcome this defect of the theory.

[55] This position seems to be consistent with the following remarks by Griliches (1971b; 8): "The time dummy approach does have the advantage, if the comparability problem can be solved, of allowing us to ignore the ever present problem of multicollinearity among the various dimensions. Using it, we may not care that in one year the coefficient of weight is high and horsepower is low while in another year these coefficients reverse themselves, as long as the two coefficients taken together hold the *joint* effect of weight and horsepower constant."

[56] But only for characteristics where we are fairly certain that more is better!

[57] These a priori beliefs can be imposed on the regression by replacing coefficients by squares of coefficients and then using nonlinear regression estimation techniques.

[58] Again there is an analogy with traditional consumer demand analysis: if we are interested in welfare analysis, we will impose the curvature conditions implied by economic theory on the econometric estimation method as is done by Diewert and Wales (1993) whereas if we are interested in testing traditional demand theory, then we would not impose these curvature conditions.

seems to me that this theory testing point of view is not what statistical agencies are interested in when they do quality adjustment: they want to make quality adjustments that are consistent with the public's a priori view that more of a desirable characteristic should increase the price of the product (or at least not decrease it).

Finally, we discuss Pakes Proposition III; namely that only *one* of the two single period hedonic regressions that pertain to the two periods under consideration needs to be used in order to undertake a quality adjustment. Obviously, this Proposition is true! However, as we have argued in the earlier sections of this paper, if two estimates are available, then it always is better to use an average of the two estimates rather than just one of them.[59] This is particularly true for Pakes' preferred index defined by the last line of (52) since this index is likely to suffer from substitution or representativity bias.[60] Hence it will usually be best to match up a Laspeyres type estimator like the estimator preferred by Pakes with a corresponding Paasche type estimator (if both are available).

10. Conclusion

The theory of hedonic regressions has left a great deal of leeway open to the empirical investigator with respect to the details of implementation of the models. Our strategy in this review of the issues has been to use some of the ideas that are present in the test approach to index number theory in an attempt to narrow down some of these somewhat arbitrary choices. The problem with arbitrary choices is that the end results may not be invariant to these choices and hence if hedonic regression techniques are used by statistical agencies, the resulting estimates of price change may not be reproducible.[61] We have made a start on narrowing down some of these choices but much work remains to be done.

In order to narrow down the range of outcomes that can result from the use of hedonic regression techniques, we make the following suggestions:

- It seems preferable to use the log of the model price as the dependent variable rather than the model price itself.

- If expenditure weights are available, use them to weight the observations as suggested in this section for a dummy variable hedonic regression and as suggested in section 3. Expenditure weights are preferable to quantity weights.

- In the context of running single period hedonic regressions, it seems preferable to run separate regressions for both periods and use a symmetric average of the results from both regressions in the final measure of price change between the two periods.

- It is preferable to use hedonic regression techniques in the context of the chain principle where the prices of period t and t−1 are compared since this will tend minimize the spread between estimates of price change over longer periods that are obtained using alternative hedonic techniques.

[59] Again, this position seems to be consistent with that of Griliches (1971b; 7).
[60] This can be most clearly seen in the matched model context where the index defined by the last line of (52) will be approximately equal to the ordinary Laspeyres index between periods s and t.
[61] The work of Heravi and Silver (2002) shows that the use of different hedonic regression techniques can lead to quite different estimates of price change.

- It seems preferable to sign restrict regression coefficients in accordance with a priori theory.

One issue that we did not discuss above is whether hedonic regressions should include brand dummy variables as independent variables. The argument against doing this is that brand dummy variables should be superfluous if we have entered all of the important characteristics of the product into the regression and hence, by including brand dummy variables, we will just use up valuable degrees of freedom and increase multicollinearity. The argument for entering brand dummy variables is that they capture in an efficient manner certain characteristics of the product that would be difficult to specify otherwise.[62] At the present stage of research in this area, I would be inclined to allow the use of brands as admissible dummy variables.

We conclude by noting that the cautious attitude towards the use of hedonic regressions expressed by Schultze and Mackie (2002) echoes the following comments made by Bean in his discussion of Court's (1939) pioneering paper on hedonic regressions:

"Mr. Court's interesting work should be carried much further, as he suggests. We should, however, not be disappointed if neither public agencies nor trade associations adopt the policy of publishing prices, values and index numbers based on the relatively tricky results that one is sure to get by applying the device of multiple correlation. The only group who would sponsor such a procedure would be the non-existent National Association of Experts in Multiple Correlation, the demand for whose services would be enormously increased." Louis H. Bean (1939; 119).

Hopefully, in the next few years, as users form a consensus on what the "best" procedures are, then the use of hedonic regressions by statistical agencies will become much more widespread and routine.[63]

[62] These difficult to specify characteristics might include reliability, availability of the product in the local marketplace and the degree of consumer knowledge about the product; i.e., some producers will choose to heavily advertise their products while others will not and the effect of this advertising may be to create a brand premium.

[63] The hedonic regression Manual being prepared by Jack Triplett (2002) should help form this consensus.

Appendix: Properties of Bilateral Weighted Hedonic Regressions

We consider some of the mathematical properties of a slight generalization of the share weighted bilateral hedonic regression model defined by (76) and (77) in section 7. The generalization is that we do not restrict the weights to sum up to 1 in each period. Thus, we replace the period s share weights s_k^s in (76) and the period t share weights s_k^t in (77) by the positive weights w_k^s and w_k^t respectively, where these weights do not necessarily sum to 1 in each period. We assume that these weight functions are known functions of the price and quantity data pertaining to periods s and t; i.e., we have for some functions, g_k^s and g_k^t:

(A1) $w_k^s = g_k^s(p^s,p^t,q^s,q^t)$ for $k \in S(s)$; $w_k^t = g_k^t(p^s,p^t,q^s,q^t)$ for $k \in S(t)$

where p^s and p^t are price vectors of the model prices for periods s and t respectively and q^s and q^t are the corresponding period s and t quantity vectors of the models sold in periods s and t. In the Propositions below, we will place further restrictions on the weighting functions g_k^s and g_k^t as they are needed.

The weighted least squares estimators for γ_{st}, β_0, β_1,...,β_N for this new model are the solutions c_{st}^*, b_0^*, b_1^*,..., b_N^* to the following quadratic weighted least squares minimization problem:[64]

(A2) $\min_{b's \text{ and } c} \{\sum_{k \in S(s)} w_k^s[\ln p_k^s - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2$
$+ \sum_{k \in S(t)} w_k^t[\ln p_k^t - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$.

The *bilateral price index* P that summarizes the overall change in prices going from period s to t is defined as the exponential of the c_{st} solution to (A2); i.e., we have:

(A3) $P(p^s,p^t,q^s,q^t) \equiv \exp[c_{st}^*]$.

We would like to show that the hedonic index number formula defined by (A3) has some of the properties that bilateral index number formulae defined over matched models usually have. Thus we are attempting to extend the test approach to index number theory[65] to weighted bilateral hedonic regressions. In particular, we would like to establish the following properties for P:

(A4) *Homogeneity of degree one in period t prices*; i.e., $P(p^s,\lambda p^t,q^s,q^t) = \lambda P(p^s,p^t,q^s,q^t)$ for all $\lambda > 0$.

(A5) *Homogeneity of degree minus one in period s prices*; i.e., $P(\lambda p^s,p^t,q^s,q^t) = \lambda^{-1}P(p^s,p^t,q^s,q^t)$ for all $\lambda > 0$.

(A6) *Identity*; i.e., if the models in the two periods are identical and the selling prices are equal so that $p^s = p^t \equiv p$ and, in addition, the same quantities of each model are sold in the two periods so that $q^s = q^t \equiv q$, then the resulting price index $P(p,p,q,q) = 1$.

[64] Throughout this Appendix, we assume that the X matrix that corresponds to the linear regression model defined by (A2) has full column rank so that the solution to (A2) exists and is unique.

[65] The test approach to index number theory was largely developed by Walsh (1901) (1921), Fisher (1911) (1922) and Eichhorn and Voeller (1976). For more recent contributions, see Diewert (1992) (1993), Balk (1995) and von Auer (2001).

(A7) *Time reversal*; i.e., $P^*(p^t,p^s,q^t,q^s) = 1/ P(p^s,p^t,q^s,q^t)$.

The above property says that if we interchange the order of our data and measure the overall change in prices going backwards from period t to s, then the resulting index $P^*(p^t,p^s,q^t,q^s)$ is equal to the reciprocal of the original index $P(p^s,p^t,q^s,q^t)$, which measured the degree of overall price change going from period s to t. In order to formally define the price index P^*, let c_{ts}^*, b_0^{**}, b_1^{**},..., b_N^{**} be the solution to the following quadratic weighted least squares minimization problem, which corresponds to reversing the ordering of the two periods:

$$\text{(A8) } \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - c_{ts} - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}.$$

The *bilateral price index* P^* that summarizes the overall change in prices going from period s to t is defined as the exponential of the c_{ts} solution to (A8); i.e., we have:

$$\text{(A9) } P^*(p^t,p^s,q^t,q^s) \equiv \exp[c_{ts}^*].$$

In the remainder of this Appendix, we shall find conditions which ensure that the tests (A4)-(A7) are satisfied.

Proposition 1: Suppose that: (i) all models are identical in the two periods so that S(s) = S(t) and $z_{kn}^s = z_{kn}^t \equiv z_{kn}$ for n = 1,...,N and k = 1,...,K; (ii) the model prices in period s are equal to the corresponding model prices in period t so that $p_k^s = p_k^t \equiv p_k$ for k = 1,...,K; (iii) the model quantities sold in period s are equal to the corresponding sales in period t so that $q_k^s = q_k^t \equiv q_k$ for k = 1,...,K; (iv) the model weights are equal across the two periods for each model so that $w_k^s = w_k^t \equiv w_k$ for k = 1,...,K. Under these hypotheses, the identity test (A6) is satisfied.

Proof: Under the above hypotheses, the least squares minimization problem (A2) becomes:

$$\text{(A10) } \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k[\ln p_k - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2 + \sum_{k\in S(t)} w_k[\ln p_k - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2\}.$$

From the general properties of minimization problems, it can be seen that the following inequality is valid:

$$\begin{aligned}\text{(A11) } &\min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k[\ln p_k - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2 + \sum_{k\in S(t)} w_k[\ln p_k - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2\} \\ \geq &\min_{b\text{'s}} \{\sum_{k\in S(s)} w_k[\ln p_k - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2\} \\ &+ \min_{b\text{'s and } c} \sum_{k\in S(t)} w_k[\ln p_k - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn})b_n]^2\}.\end{aligned}$$

Let b_0^*, b_1^*,..., b_N^* solve the first minimization problem on the right hand side of (A11). Now look at the second minimization problem on the right hand side of (A11). Obviously the parameters c_{st} and b_0 cannot be separately identified so one of them can be set equal to zero; we choose to set $c_{st} = 0$. But after setting $c_{st} = 0$, we see that the second minimization problem is identical to the first minimization problem on the right hand side of (A11), and hence $c_{st}^* = 0$ and b_0^*, b_1^*,..., b_N^* solve the second minimization problem. However, $c_{st}^* = 0$ and b_0^*, b_1^*,..., b_N^* are feasible for the minimization problem on the left hand side of (A11) and since the objective function evaluated at this feasible solution attains a lower bound, we conclude

that $c_{st}^* = 0$ and $b_0^*, b_1^*,..., b_N^*$ solves (A10). But $c_{st}^* = 0$ implies $P(p,p,q,q) \equiv \exp[c_{st}^*] = \exp[0] = 1$, which is the desired result (A6). Q.E.D.

Proposition 2: Suppose that the weight functions defined by (A1) are homogeneous of degree zero in the components of the period t price vector p^t, so that for all $\lambda > 0$, $g_k^s(p^s,\lambda p^t,q^s,q^t) = g_k^s(p^s,p^t,q^s,q^t)$ for $k \in S(s)$ and $g_k^t(p^s,\lambda p^t,q^s,q^t) = g_k^t(p^s,p^t,q^s,q^t)$ for $k \in S(t)$. Then the hedonic price index $P(p^s,p^t,q^s,q^t)$ defined by (A3) will satisfy the homogeneity of degree one property (A4).[66]

Proof: Let $c_{st}^*, b_0^*, b_1^*,..., b_N^*$ solve the initial minimization problem (A2) before we multiply the period t price vector by $\lambda > 0$. Now consider a new weighted least squares minimization problem where p^t has been replaced by λp^t. Under our hypotheses, the weights will not be changed by this change in the period t prices and so the new minimization problem will be:

$$\text{(A12)} \quad \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t + \ln \lambda - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}.$$

$$\text{(A13)} \quad = \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - c_{st}' - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$$

where the new c_{st} variable is defined as follows:

$$\text{(A14)} \quad c_{st}' \equiv c_{st} - \ln \lambda.$$

Denote the solution to (A13) as $c_{st}^{**'}, b_0^{**}, b_1^{**},..., b_N^{**}$. However, it can be seen that the solution to (A13) is exactly the same as the solution to the initial problem, (A2). Hence $c_{st}^{**'} = c_{st}^*$, and the c_{st} solution to (A12), which we denote by c_{st}^{**}, satisfies (A14):

$$\text{(A15)} \quad c_{st}^* = c_{st}^{**'} = c_{st}^{**} - \ln \lambda \text{ or}$$
$$\text{(A16)} \quad c_{st}^{**} = c_{st}^* + \ln \lambda.$$

Hence

$$\begin{aligned} \text{(A17)} \quad P(p^s,\lambda p^t,q^s,q^t) &\equiv \exp[c_{st}^{**}] \\ &= \exp[c_{st}^* + \ln \lambda] && \text{using (A16)} \\ &= \lambda \exp[c_{st}^*] \\ &= \lambda P(p^s,p^t,q^s,q^t) && \text{using definition (A3)} \end{aligned}$$

which establishes the desired result (A4). Q.E.D.

Proposition 3: Suppose that the weight functions defined by (A1) are homogeneous of degree zero in the components of the period s price vector p^s, so that for all $\lambda > 0$, $g_k^s(\lambda p^s,p^t,q^s,q^t) = g_k^s(p^s,p^t,q^s,q^t)$ for $k \in S(s)$ and $g_k^t(\lambda p^s,p^t,q^s,q^t) = g_k^t(p^s,p^t,q^s,q^t)$ for $k \in S(t)$. Then the hedonic price index $P(p^s,p^t,q^s,q^t)$ defined by (A3) will satisfy the homogeneity of degree minus one property (A5).[67]

[66] We assume that the period t quantity vector q^t remains the same if the period t prices change from p^t to λp^t.
[67] We assume that the period s quantity vector q^s remains the same if the period s prices change from p^s to λp^s.

Proof: Let c_{st}^*, b_0^*, b_1^*,..., b_N^* solve the initial minimization problem (A2) before we multiply the period s price vector by $\lambda > 0$. Now consider a new weighted least squares minimization problem where p^s has been replaced by λp^s. Under our hypotheses, the weights will not be changed by this change in the period s prices and so the new minimization problem will be:

$$\text{(A18)} \quad \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s + \ln \lambda - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - c_{st} - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}.$$

$$\text{(A19)} \quad = \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - b_0' - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - c_{st}' - b_0' - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$$

where the new b_0 and c_{st} variables are defined as follows:

$$\text{(A20)} \quad b_0' \equiv b_0 - \ln \lambda ; \; c_{st}' \equiv c_{st} + \ln \lambda.$$

Denote the solution to (A19) as $c_{st}^{**\prime}$, $b_0^{**\prime}$, b_1^{**},..., b_N^{**}. However, it can be seen that the solution to (A19) is exactly the same as the solution to the initial problem, (A2). Hence $c_{st}^{**\prime} = c_{st}^*$ and $b_0^{**\prime} = b_0^*$. Thus the c_{st} solution to (A18), which we denote by c_{st}^{**}, satisfies the following equations, where we have substituted into equations (A20):

$$\text{(A21)} \quad b_0^* \equiv b_0^{**} - \ln \lambda ; \; c_{st}^* \equiv c_{st}^{**} + \ln \lambda.$$

Using the second equation in (A21), we have:

$$\text{(A22)} \quad c_{st}^{**} = c_{st}^* - \ln \lambda.$$

Hence

$$\begin{aligned} \text{(A23)} \quad P(\lambda p^s, p^t, q^s, q^t) &\equiv \exp[c_{st}^{**}] \\ &= \exp[c_{st}^* - \ln \lambda] && \text{using (A22)} \\ &= \lambda^{-1} \exp[c_{st}^*] \\ &= \lambda^{-1} P(p^s, p^t, q^s, q^t) && \text{using definition (A3)} \end{aligned}$$

which establishes the desired result (A5). Q.E.D.

Note that in both Propositions 2 and 3, it is not necessary that the weights w_k^s and w_k^t sum to one for each period s and t.

Proposition 4: The bilateral hedonic price index which measures price change going from period s to t, $P(p^s,p^t,q^s,q^t)$ defined by (A3), and the bilateral hedonic price index which measures price change going from period t to s, $P^*(p^t,p^s,q^t,q^s)$ defined by (A9), satisfy the time reversal test (A7).

Proof: As usual, denote the solution to (A2) as c_{st}^*, b_0^*, b_1^*,..., b_N^*. The minimization problem, which corresponds to reversing the ordering of the two periods, is (A24) below and it has the solution c_{ts}^*, b_0^{**}, b_1^{**},..., b_N^{**}:

$$\text{(A24)} \quad \min_{b\text{'s and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - c_{ts} - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$$

$$(A25) = \min_{b's \text{ and } c} \{\sum_{k\in S(s)} w_k^s[\ln p_k^s - b_0' - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2 + \sum_{k\in S(t)} w_k^t[\ln p_k^t - c_{ts}' - b_0' - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$$

where we have defined the new variables b_0' and c_{ts}' in terms of the old variables b_0 and c_{ts} as follows:

$$(A26)\ b_0' \equiv b_0 + c_{ts}\ ;\ c_{st}' \equiv -c_{st}\ .$$

Denote the solution to (A25) as $c_{ts}^{*\prime}$, $b_0^{**\prime}$, $b_1^{**\prime}$,..., $b_N^{**\prime}$. However, it can be seen that the solution to (A25) is exactly the same as the solution to the initial problem, (A2). Hence $c_{ts}^{*\prime} = c_{st}^*$ and $b_0^{**\prime} = b_0^*$. Thus the c_{ts} solution to (A24), which we denoted by c_{ts}^*, satisfies the following equations, where we have substituted into equations (A26):

$$(A27)\ b_0^* \equiv b_0^{**} + c_{ts}^*\ ;\ c_{st}^* \equiv -c_{ts}^*.$$

Using definition (A9), we have:

$$\begin{aligned}(A28)\ P^*(p^t,p^s,q^t,q^s) &\equiv \exp[c_{ts}^*] \\ &= \exp[-c_{st}^*] && \text{using (A27)} \\ &= 1/\exp[c_{st}^*] \\ &= 1/P(p^s,p^t,q^s,q^t) && \text{using definition (A3)}\end{aligned}$$

which establishes the desired result (A7). Q.E.D.

Proposition 5: Let c_{st}^*, b_0^*, b_1^*,..., b_N^* denote the solution to the weighted least squares problem (A2). Then c_{st}^*, which is the logarithm of the bilateral hedonic price index $P(p^s,p^t,q^s,q^t)$ defined by (A3), satisfies the following equation:[68]

$$\begin{aligned}(A29)\ [\sum_{k\in S(t)} w_k^t]\, c_{st}^* &= \sum_{k\in S(t)} w_k^t \ln p_k^t - \sum_{k\in S(s)} w_k^s \ln p_k^s - [\sum_{k\in S(t)} w_k^t]\, b_0^* \\ &\quad + [\sum_{k\in S(s)} w_k^s]\, b_0^* - \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t)b_n^* + \sum_{k\in S(s)} w_k^s \sum_{n=1}^N f_n(z_{kn}^s)b_n^* \\ &= \sum_{k\in S(t)} w_k^t [\ln p_k^t - b_0^* - \sum_{n=1}^N f_n(z_{kn}^t)b_n^*] - \sum_{k\in S(s)} w_k^s [\ln p_k^s - b_0^* - \sum_{n=1}^N f_n(z_{kn}^t)b_n^*].\end{aligned}$$

Proof: The solution c_{st}^*, b_0^*, b_1^*,..., b_N^* to the minimization problem (A2) can be obtained by applying least squares to the following linear regression model:

$$\begin{aligned}(A30)\ (w_k^s)^{1/2} \ln p_k^s &= (w_k^s)^{1/2}[b_0^* + \sum_{n=1}^N f_n(z_{kn}^s)b_n^*] + e_k^s\ ; && k\in S(s); \\ (w_k^t)^{1/2} \ln p_k^t &= (w_k^t)^{1/2}[c_{st}^* + b_0^* + \sum_{n=1}^N f_n(z_{kn}^t)b_n^*] + e_k^t\ ; && k\in S(t).\end{aligned}$$

We have inserted the optimal least squares estimators, c_{st}^*, b_0^*, b_1^*,..., b_N^*, into equations (A30) so that we can use these equations to define the least squares residuals e_k^s and e_k^t for the period s and t observations. It is well known that the column vector of these residuals is orthogonal to the columns of the X matrix, which correspond to the exogenous variables on the right hand side of equations (A30). These orthogonality relations applied to the columns that correspond to the constant term b_0 and the time dummy variable c_{st} give us the following 2 equations:

[68] The two equations in (A29) are generalizations of a similar formula derived by Triplett and McDonald (1977; 150) in the unweighted context. This unweighted formula was also used by Triplett (2000; 39). The technique of proof used in this Proposition was used in section 4 of Diewert (2001).

(A31) $0 = \sum_{k\in S(s)} w_k^s \ln p_k^s + \sum_{k\in S(t)} w_k^t \ln p_k^t - [\sum_{k\in S(t)} w_k^t] c_{st}^* - [\sum_{k\in S(s)} w_k^s] b_0^*$
$- [\sum_{k\in S(t)} w_k^t] b_0^* - \sum_{k\in S(s)} w_k^s \sum_{n=1}^N f_n(z_{kn}^s) b_n^* - \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t) b_n^*$;

(A32) $0 = \sum_{k\in S(t)} w_k^t \ln p_k^t - \sum_{k\in S(t)} w_k^t c_{st}^* - \sum_{k\in S(t)} w_k^t b_0^* - \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t) b_n^*$.

Equation (A32) can be rewritten as:

(A33) $\sum_{k\in S(t)} w_k^t \ln p_k^t = [\sum_{k\in S(t)} w_k^t] c_{st}^* + [\sum_{k\in S(t)} w_k^t] b_0^* + \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t) b_n^*$.

Subtracting (A32) from (A31) leads to the following equation:

(A34) $\sum_{k\in S(s)} w_k^s \ln p_k^s = [\sum_{k\in S(s)} w_k^s] b_0^* + \sum_{k\in S(s)} w_k^s \sum_{n=1}^N f_n(z_{kn}^s) b_n^*$.

Finally, subtracting (A34) from (A33) leads to (A29) after a bit of rearrangement. Q.E.D.

Corollary 5.1: If the models are identical during the two periods and the weights are also identical across periods for the same model, then the hedonic price index $P(p^s,p^t,q^s,q^t)$ defined by (A3) is equal to a weighted geometric mean of the model price relatives, where the weights are proportional to the common model weights, $w_k^s = w_k^t \equiv w_k$.

Proof: Under the stated hypotheses, the last 4 sets of terms on the right hand side of (A29) sum to zero and hence the logarithm of $P(p^s,p^t,q^s,q^t)$ is equal to:

(A35) $c_{st}^* = \sum_{k=1}^K w_k \ln [p_k^t/p_k^s]/[\sum_{j=1}^K w_j]$

which establishes the desired result. Q.E.D.

Corollary 5.2: If the models are identical during the two periods and the weight for model k is chosen to be the arithmetic average of the expenditure shares on the model for the two periods, $(1/2)s_k^s + (1/2)s_k^t$, then the hedonic price index $P(p^s,p^t,q^s,q^t)$ defined by (A3) is equal to the Törnqvist (1936) price index.

Proof: Apply (A35) with $w_k \equiv (1/2)s_k^s + (1/2)s_k^t$. Q.E.D.

We conclude this Appendix by noting that the second equation in (A29) has a nice interpretation in the light of our discussion of quality adjusted price relatives in section 4 above: it can be seen that c_{st}^* is equal to a weighted *sum* of the logarithms of the *quality adjusted prices* of the models sold in period t less another weighted *sum* of the logarithms of the *quality adjusted prices* of the models sold in period s. If the weights sum to unity in each period, then the two weighted sums become weighted *averages* of the logarithms of quality adjusted prices.

References

Balk, B. M. (1995), "Axiomatic Price Index Theory: A Survey", *International Statistical Review* 63:1, 69-93.

Bean, L. H. (1939), "Discussion of Mr. Court's Paper on Hedonic Price Indexes", pp. 118-119 in *The Dynamics of Automobile Demand*, New York: General Motors Corporation.

Berndt, E. R., Z. Griliches and N. J. Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990's", *Journal of Econometrics* 68, 243-268.

Chatterjee, S. and A. S. Hadi (1988), *Sensitivity Analysis in Linear Regression*, New York: John Wiley.

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

Diewert, W. E. (1987), "Index Numbers", pp. 767-780 in *The New Palgrave: A Dictionary of Economics, Volume 2*, J. Eatwell, M. Milgate and P. Newman (eds.), London: Macmillan; reprinted as pp. 71-104 in W. E. Diewert and A. O. Nakamura (eds.) (1993), *Essays in Index Number Theory, Volume 1*, Amsterdam: North-Holland.

Diewert, W. E. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited", *Journal of Productivity Analysis* 3, 211-248; reprinted as pp. 317-353 in *Essays in Index Number Theory, Volume 1*, W. E. Diewert and A. O. Nakamura (eds.), Amsterdam: North-Holland, 1993.

Diewert, W. E. (1993), "The Early History of Price Index Research", pp. 33-65 in *Essays in Index Number Theory, Volume 1*, W. E. Diewert and A. O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W. E. (1997), "Commentary", *Federal Reserve Bank of St. Louis Review* 79:3 (May/June), 127-138.

Diewert, W. E. (2001), "Hedonic Regressions: A Consumer Theory Approach", Discussion Paper 01-12, Department of Economics, University of British Columbia, Vancouver Canada, V6T 1Z1.

Diewert, W. E. (2002a), "Harmonized Indexes of Consumer Prices: Their Conceptual Foundations", European Central Bank Working Paper Series No. 130, Frankfurt: ECB.

Diewert, W. E. (2002b), "Weighted Hedonic Regressions and Index Number Formulae", paper tabled at the Third Annual Joint NBER and CRIW Measurement Workshop, Cambridge MA, July 29-31, 2002.

Diewert, W. E. and T. J. Wales (1993), "Linear and Quadratic Spline Models For Consumer Demand Functions", *Canadian Journal of Economics* 26, 77-106.

Dhrymes, P. J. (1971), "Price and Quality Changes in Consumer Capital Goods: An Empirical Study", pp. 88-149 in *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press.

Eichhorn, W. and J. Voeller (1976), *Theory of the Price Index: Fisher's Test Approach and Generalizations*, Lecture Notes in Economics and Mathematical Systems, Volume 140, Berlin: Springer-Verlag.

Fisher, I. (1911), *The Purchasing Power of Money*, London: Macmillan.

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

Greene, W. H. (1993), *Econometric Analysis*, Second Edition, Englewood Cliffs, New Jersey: Prentice Hall.

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

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

Heravi, S. and M. Silver (2002), "On the Stability of Hedonic Coefficients and their Implications for Quality Adjusted Price Change Measurement", paper presented at the Third Annual Joint NBER and CRIW Measurement Workshop, Cambridge MA, July 29-31, 2002.

Hicks, J. R. (1940), "The Valuation of the Social Income", *Economica* 7, 105-140.

Hulten, C. R. (2002), "Price Hedonics: A Critical Review", paper presented at the Third Annual Joint NBER and CRIW Measurement Workshop, Cambridge MA, July 29-31, 2002.

Konüs, A. A. (1924), "The Problem of the True Index of the Cost of Living", translated in *Econometrica* 7, (1939), 10-29.

Muellbauer, J. (1974), "Household Production Theory, Quality and the 'Hedonic Technique'", *The American Economic Review* 64:6, 977-994.

Pakes, A. (2001), "A Reconsideration of Hedonic Price Indices with an Application to PC's", NBER Working Paper 8715, Cambridge MA, revised November 2001.

Pollak, R. A. (1983), "The Treatment of 'Quality' in the Cost of Living Index", *Journal of Public Economics* 20, 25-53.

Rao, D. S. Prasada (2002), "On the Equivalence of Weighted country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons", School of Economics, University of New England, Armidale, Australia, March.

Silver, M. and Heravi, S. (2001), "Scanner Data and the Measurement of Inflation", *The Economic Journal*, 111 (June), 384-405.

Silver, M. and Heravi, S. (2002a), "The Measurement of Quality-Adjusted Price Changes", Mathew Shapiro and Rob Feenstra (eds.), *Scanner Data and Price Indexes*, National Bureau of Economic Research, Studies in Income and Wealth, vol. 61, Chicago: University of Chicago Press.

Silver, M. and Heravi, S. (2002b), "Why the CPI Matched Models Method May Fail Us: Results from an Hedonic and Matched Experiment Using Scanner Data", European Central Bank Working Paper No. 144, Frankfurt: ECB.

Schultze, C. L. and C. Mackie (eds.) (2002), *At What Price? Conceptualizing and Measuring Cost-of-Living Indexes*, Washington D. C.: National Academy Press.

Theil, H. (1967), *Economics and Information Theory*, Amsterdam: North-Holland.

Törnqvist, Leo (1936), "The Bank of Finland's Consumption Price Index", *Bank of Finland Monthly Bulletin* 10: 1-8.

Triplett, J. E. (1983), "Concepts of Quality in Input and Output Price Measures: A Resolution of the User Value and Resource Cost Debate", pp. 269-311 in *The U. S. National Income and Product Accounts: Selected Topics*, NBER Studies in Income and Wealth Volume 47, M. F. Foss (ed.), Chicago: The University of Chicago Press.

Triplett, J. D. (2000), *Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products*, November 10 draft, Paris: OECD.

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

von Auer, L. (2001), "An Axiomatic Checkup for Price Indices", working Paper No. 1/2001, Faculty of Economics and Management, Otto von Guericke University Magdeburg, Postfach 4120, 39016 Magdeburg, Germany.

Walsh, C. M. (1901), *The Measurement of General Exchange Value*, New York: Macmillan and Co.

Walsh, C. M. (1921), *The Problem of Estimation*, London: P.S. King & Son.

Whittaker, E. T. and G. Robinson (1940), *The Calculus of Observations*, Third Edition, London: Blackie & Sons.