ADJACENT PERIOD DUMMY VARIABLE HEDONIC REGRESSIONS

AND BILATERAL INDEX NUMBER THEORY

by

Erwin Diewert
Department of Economics
University of British Columbia

August 2005

Discussion Paper No.: 05-11

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, CANADA V6T 1Z1

http://www.econ.ubc.ca

Adjacent Period Dummy Variable Hedonic Regressions and Bilateral Index Number Theory

Erwin Diewert[1], August 5, 2005.
Discussion Paper No. 05-11,
Department of Economics,
University of British Columbia,
Vancouver, B.C.,
Canada, V6T 1Z1.
email: diewert@econ.ubc.ca

Abstract

A hedonic regression regresses the price of various models of a product (or service) on the characteristics that describe the product. The issue that is addressed in this paper is the following one: if information on the prices and quantities purchased of the various models is available, then how should this extra information be used? The paper suggests various methods of weighting that might be used in an adjacent period dummy time variable hedonic regression framework. Some of the ideas that are present in the test approach to index number theory are used in an attempt to cast some light on the consequences of different types of weighting. Section 2 weighting in a single period hedonic regression framework. Section 3 discusses the adjacent period hedonic regression model without weighting. Section 4 introduces weighting into the adjacent period framework and the ideas developed in the earlier sections are applied. Section 5 specializes the general model of section 4 to the "pure" dummy variable approach to the specification of characteristics of models and section 6 concludes.

Key Words

Hedonic regression, test approach to index number theory, consumer theory, characteristics, quality change, matched models, consumer price index.

Journal of Economic Literature Classification System Numbers

C23, C43, C51, D11, D12, E31

[1] The author is indebted to Ana Aizcorbe, Ernst Berndt, Carol Corrado, Melvyn Fuss, Robert Gordon, Jacques Mairesse, Alice Nakamura, Mick Silver and Manuel Trajtenberg for helpful comments and to the SSHRC of Canada for financial support. None of the above are responsible for the views expressed in this chapter. Sections 2-4 and the Appendix in this paper originally appeared in Diewert (2003b). This paper will appear in *Essays in honor of Zvi Griliches*, J. Mairesse and M. Trajtenberg (eds.), forthcoming in a Special issue of the *Annales d'Economie et Statistique*, 2005.

1. Introduction

Some recent publications have revived interest in the topic of hedonic regressions. The first 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 second 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] Finally, the comprehensive monograph by Triplett (2004) argued strongly in favor of the use of hedonic regression methods to adjust prices for quality changes over other methods that have been suggested.

One important problem area associated with the use of hedonic regressions is the issue of *whether the regressions should be weighted or not*. This is the issue that we will address in this chapter.[3] Our approach to answering this question will be somewhat novel: we will try to use ideas that occur in the index number literature on weighting to help guide us in evaluating alternative approaches to the weighting question in the hedonic regression context. Thus consider the problem of constructing a Consumer Price Index that compares prices between two periods. Index number theory approaches to this problem end up constructing a CPI between the two periods as a share weighted average of the price relatives for all the commodities in the domain of definition of the index that can be matched between the two periods. Hedonic regressions are used to extend this framework to situations where the quality of the commodities may change over time (and so complete matching across the two periods is not possible) but the overall goal is the same as with a standard CPI: we want a single number that describes the "average" price change between the two periods. Thus in this chapter, we will focus on the *dummy variable adjacent year hedonic regression technique* initially suggested by Court (1939; 109-111) and used by Berndt, Griliches and Rappaport (1995; 260) and many others, since this hedonic regression framework gives us an unambiguous measure of price change between the two periods under consideration.[4]

Our general strategy will be to suggest various methods of weighting that might be used in this adjacent period dummy time variable hedonic regression framework, look at the resulting measures of overall price change but then specialize the model to the matched model context so that the resulting measure of price change can be compared to more traditional index number measures of overall price change. Thus we set the stage for this methodology by first discussing weighting in a single period hedonic regression framework in section 2 and then in section 3, we discuss the adjacent period hedonic regression model without weighting. In section 4, we introduce weighting into the adjacent period framework and apply our index number type methodology. Section 5

---

[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] The only extensive discussion of this issue that I am aware of is by Triplett (2004; 189-194).

[4] When we estimate separate hedonic regressions for both periods, there are many ways that this information from the two regressions can be used in order to obtain a single overall measure of price change; see Silver and Heravi (2002) (2003) (2004) (2005) and Triplett (2004) for discussions and comparisons of these various methods.

specializes the general model of section 4 to an interesting case considered by Aizcorbe, Corrado and Doms (2000), who introduced a "pure" dummy variable approach to the specification of characteristics of models in addition to the usual time dummies. Section 6 concludes and an Appendix presents proofs of the various results that are used in sections 3-5.

2. Quantity Weights versus Expenditure Weights

As an introduction to our main topic, we discuss alternative methods of weighting model prices in a single equation hedonic regression. Thus if information on model prices, characteristics and sales to households is available to a statistical agency producing a Consumer Price Index, then how exactly could the extra information on sales be used in running a single period hedonic regression?

We introduce some notation at this point. We suppose that price data have been collected on K models or varieties of a commodity for some period t.[5] Thus p_k^t is the price of model k in period t and $k \in S(t)$ where S(t) is the set of models that are actually purchased by households in period t. For $k \in S(t)$, denote the number of these type k models sold during period t by q_k^t.[6] 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 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 $k \in S(t)$. We shall consider only linear hedonic regressions in this chapter. Hence, the *unweighted linear hedonic regression for period t* has the following form:[7]

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

where the β_n^t are unknown parameters to be estimated, the ε_k^t are independently distributed error terms with mean 0 and variance σ^2 and the functions f and f_n are known functions that are used to transform the data (typically, they 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).

Usually, econometric 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 an idea taken from index number theory—namely 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

[5] Models purchased in different outlets can be regarded as separate varieties or not, depending on the context.

[6] If a particular model k is purchased 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.

[7] Note that the linear regression model defined by (1) can only provide a first order approximation to a general hedonic function. Diewert (2003a) made a case for considering second order approximations but in this chapter, we will follow current practice and consider only linear approximations.

the period t regression is *representative* of the sales that actually occurred during the period.[8]

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 :

$$
\begin{aligned}
(2)\ f(p_1^t) &= \beta_0^t + f_1(z_{11}^t)\beta_1^t + \varepsilon_1^t ;\\
f(p_2^t) &= \beta_0^t + f_1(z_{21}^t)\beta_1^t + \varepsilon_2^t ;\\
f(p_3^t) &= \beta_0^t + f_1(z_{31}^t)\beta_1^t + \varepsilon_3^t .
\end{aligned}
$$

Note that each of the 3 observations gets an equal weight in the period t hedonic regression model defined by (2). 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 in the regression 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:

$$
\begin{aligned}
(3)\ 1_1 f(p_1^t) &= 1_1\beta_0^t + 1_1 f_1(z_{11}^t)\beta_1^t ;\\
1_2 f(p_2^t) &= 1_2\beta_0^t + 1_2 f_1(z_{21}^t)\beta_1^t ;\\
1_3 f(p_3^t) &= 1_3\beta_0^t + 1_3 f_1(z_{31}^t)\beta_1^t
\end{aligned}
$$

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 (2):

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

Comparing (2) and (4), it can be seen that the observations in (4) are equal to the corresponding observations in (2), except that the dependent and independent variables in

[8] Thus our representative approach follows along the lines of Theil's (1967; 136-138) weighted stochastic approach to index number theory, which is also pursued by Clements and Izan (1981), Selvanathan and Rao (1994), Rao (2002) and Diewert (1995) (2004). 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.

observation k of (2) 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 (4). A sampling framework for (4) 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 (3) 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 (4). Then it is straightforward to show that these two sets of least squares estimators are the same[9]; i.e., we have:

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

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 (3) is to obtain the least squares estimators for the transformed model (4). This equivalence between the two models provides a justification for using the weighted model (4) in place of the original model (2). The advantage in using the transformed model (4) over the "representative" model (3) is that we can develop a sampling framework for (4) but not for (3), since the (omitted) error terms in (3) cannot be assumed to be distributed independently of each other.[10] However, in view of the equivalence between the least squares estimators for models (3) and (4), we can now be comfortable that the regression model (4) weights observations according to their quantitative importance in period t. Hence if we take the point of view that regards weighting according to economic importance as fundamental, then we can recommend the use of the weighted hedonic regression model (4) over its unweighted counterpart (2).

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:

(6) $v_k^t \equiv p_k^t q_k^t$; $\qquad\qquad k\in S(t)$.

Now consider again the simple unweighted hedonic regression model defined by (2) 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 (2) 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):

---

[9] 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).

[10] It is possible to develop a descriptive statistics interpretation for b_0^t and b_1^t, the least squares estimators for the β_0^t and β_1^t parameters in (3); see section 8 in Diewert (2004).

$$(7)\ 1_{1*}f(p_1{}^t) = 1_{1*}\beta_0{}^t + 1_{1*}f_1(z_{11}{}^t)\beta_1{}^t\ ;$$
$$1_{2*}f(p_2{}^t) = 1_{2*}\beta_0{}^t + 1_{2*}f_1(z_{21}{}^t)\beta_1{}^t\ ;$$
$$1_{3*}f(p_3{}^t) = 1_{3*}\beta_0{}^t + 1_{3*}f_1(z_{31}{}^t)\beta_1{}^t\ .$$

Now consider the following value transformation of the original unweighted hedonic regression model (2):

$$(8)\ (v_1{}^t)^{1/2}\ f(p_1{}^t) = (v_1{}^t)^{1/2}\ \beta_0{}^t + (v_1{}^t)^{1/2}\ f_1(z_{11}{}^t)\beta_1{}^t + \varepsilon_1{}^{t**}\ ;$$
$$(v_2{}^t)^{1/2}\ f(p_2{}^t) = (v_2{}^t)^{1/2}\ \beta_0{}^t + (v_2{}^t)^{1/2}\ f_1(z_{21}{}^t)\beta_1{}^t + \varepsilon_2{}^{t**}\ ;$$
$$(v_3{}^t)^{1/2}\ f(p_3{}^t) = (v_3{}^t)^{1/2}\ \beta_0{}^t + (v_3{}^t)^{1/2}\ f_1(z_{31}{}^t)\beta_1{}^t + \varepsilon_3{}^{t**}\ .$$

Comparing (2) and (8), it can be seen that the observations in (8) are equal to the corresponding observations in (2), except that the dependent and independent variables in observation k of (2) 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 left hand side variables in (8). Again, a sampling framework for (8) is available if we assume that the transformed residuals $\varepsilon_k{}^{t**}$ 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 $\beta_0{}^t$ and $\beta_1{}^t$ in (7) and (8) are the same. Thus a shortcut method for obtaining the least squares estimators for the unknown parameters, $\beta_0{}^t$ and $\beta_1{}^t$, which occur in the value weights representative model (7) is to obtain the least squares estimators for the transformed model (8). This equivalence between the two models provides a justification for using the value weighted model (8) in place of the value weights representative model (7). As before, the advantage in using the transformed model (8) over the value weights representative model (7) is that we can develop a sampling framework for (8) but not for (7), since the (omitted) error terms in (7) cannot be assumed to be distributed independently of each other.

From the viewpoint of index number theory, it seems to us that the quantity weighted and value weighted models are clear improvements over the original unweighted model (2). 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.[11] In the

[11] "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).

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 in the single period context, it appears to us that value weighting is clearly preferable. Thus we are taking the point of view that the main purpose of a period t hedonic regression is to enable 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.

We will draw on the material in this section in section 4 below. However, in the following section, we provide an introduction to the theory of weighted adjacent period hedonic regressions by considering the unweighted case first.

3. 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:

$$(9)\quad f(p_k^s) = \beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^s)\beta_n + \varepsilon_k^s\,; \qquad k\in S(s);$$
$$(10)\quad f(p_k^t) = \gamma_{st} + \beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^t)\beta_n + \varepsilon_k^t\,; \qquad k\in S(t);$$

where the variables in (9) and (10) are defined in the same manner as in equation (1) above. In particular, ε_k^s and ε_k^t are independently distributed error terms with mean 0 and variance σ^2. Note that the β regression coefficients in (9) are constrained to be the same as the corresponding β coefficients in (10). Note also that equations (10) 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.[12]

[12] 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.

Before proceeding further, we briefly discuss some of the advantages and disadvantages of the dummy variable model defined by (9) and (10) 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.[13] The main advantages of the dummy variable method are that it conserves degrees of freedom and is less subject to multicollinearity problems[14] and there is no ambiguity about the measure of overall price change between periods s and t.[15]

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. However, the bilateral model defined by (9) and (10) 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.[16] Let us estimate the unknown parameters in (9) and (10) 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 (9) and (10) 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:

(11) $p_k^s = b_0 + \sum_{n=1}^N f_n(z_{kn}^s)b_n + e_k^s$; $\qquad k \in S(s)$;

[13] For treatments of the issues involved in averaging the results of two period specific hedonic regressions to obtain measures of overall price change, see Koskimäki and Vartia (2001) for the unweighted case and Silver and Heravi (2002) (2003) (2004) (2005), Diewert (2003b) and Triplett (2004) for the weighted case.

[14] 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."

[15] 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."

[16] This model is not without interest. Suppose there is a minimal base period model but various additional amounts of useful characteristics can be purchased at constant prices in each period. Then it would be natural to set the f_n to be equal to identity functions as well as f and within each period, the linear functional form for the hedonic regression would be quite appropriate.

$$(12)\ p_k^{\,t} = c_{st} + b_0 + \sum_{n=1}^{N} f_n(z_{kn}^{\,t})b_n + e_k^{\,t}\ ; \qquad k \in S(t).$$

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:[17]

$$(13)\ c_{st} + b_0 + \sum_{n=1}^{N} f_n(z_{kn})b_n = \lambda[\ b_0 + \sum_{n=1}^{N} f_n(z_{kn})b_n]\ ; \qquad k = 1,\ldots,K.$$

In general, if $K > N+2$ and $\lambda \neq 1$, it can be seen that equations (13) cannot be solved for any coefficients c_{st}, b_0, $b_1,\ldots,b_N$. Hence, *our conclusion is that the linear time dummy hedonic regression model defined by (11) and (12) 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.[18] Of course, this homogeneity problem with the linear dummy variable regression model can be solved if we replace equations (12) by the following equations:

$$(14)\ p_k^{\,t} = c_{st}[b_0 + \sum_{n=1}^{N} f_n(z_{kn}^{\,t})b_n] + e_k^{\,t}\ ; \qquad k \in S(t).$$

In equations (14), the time dummy variable, c_{st}, now appears in a multiplicative fashion. Thus, the problem with the estimating equations (11) and (14) is that *we no longer have a linear regression model*; nonlinear estimation techniques would have to be used.

This is our first example of how the test approach that is commonly used in bilateral index number theory can be adapted to the adjacent period or time dummy hedonic regression context in order to obtain useful restrictions on the form of the hedonic regression. Many additional examples will be presented in what follows.[19]

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 (11) and (12) are the following equations:

$$(15)\ \ln p_k^{\,s} = b_0 + \sum_{n=1}^{N} f_n(z_{kn}^{\,s})b_n + e_k^{\,s}\ ; \qquad k \in S(s);$$
$$(16)\ \ln p_k^{\,t} = c_{st} + b_0 + \sum_{n=1}^{N} f_n(z_{kn}^{\,t})b_n + e_k^{\,t}\ ; \qquad k \in S(t).$$

[17] Let $z_{kn} \equiv z_{kn}^{\,s} = z_k^{\,t}$ denote the common amount of characteristic n that the identical model k has in each period.
[18] Diewert (2003a) also argued on theoretical grounds that dummy variable hedonic regression models that used untransformed prices as dependent variables did not have good properties.
[19] The test approach to bilateral index number theory is reviewed in Diewert (1992), Balk (1995) and the ILO (2004; 292-299).

Exponentiating both sides of (15) and (16) leads to the following equations that will be satisfied by the data and the least squares estimators for (15) and (16):

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

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:

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

It can be seen that if we choose $c_{st} = \ln \lambda$, then we can satisfy equations (19). Hence we conclude (from a test approach perspective) that if we want to use linear regression techniques to estimate the parameters of the hedonic regression, then *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*.[20]

The bilateral log hedonic regression model is defined by (9) and (10) 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 a sample estimator of this population measure is:

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

where c_{st} is the least squares estimator for the shift parameter γ_{st}. Note that we put the shift parameter in equations (10) rather than in equations (9). 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:

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

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 (21) and (22), 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:

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

[20] However, recall our earlier qualification which noted that if additional amounts of all characteristics can be purchased at constant prices in each period, then a nonlinear regression model with the f and f_n set equal to identity functions is preferable.

The question now is: how does P(s,t) defined by (20) relate to P(t,s) defined by (23)? Ideally, we would like these two estimators of price change to satisfy the following *time reversal test*:

(24) P(t,s) =1/P(s,t).

If we compare the original log linear regression model defined by (9) and (10) (with f being the log transformation) with the new model defined by (21) and (22), 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 (21) and (22) while γ_{st} appears in the second set of equations in (9) and (10). The transpose of the column in the X matrix that corresponds to γ_{ts} in (21) and (22) 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 (9) and (10) 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 (9) and (10) is equal to the subspace spanned by the X columns corresponding to β_0^* and γ_{ts} in (21) and (22) and the two sets of parameters are related by the following equations:

(25) $[0_1^T,1_2^T]\,\gamma_{st} + [1_1^T,1_2^T]\,\beta_0 = [1_1^T,0_2^T]\,\gamma_{ts} + [1_1^T,1_2^T]\,\beta_0^*$.

Equations (25) are equivalent to the following 2 equations in the four variables γ_{st}, β_0, γ_{ts} and β_0^*:

(26) $0\,\gamma_{st} + 1\,\beta_0 = 1\,\gamma_{ts} + 1\,\beta_0^*$;
$1\,\gamma_{st} + 1\,\beta_0 = 0\,\gamma_{ts} + 1\,\beta_0^*$.

Thus given γ_{st} and β_0, the corresponding γ_{ts} and β_0^* can be obtained using equations (26) as:

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

Equations (27) also hold for the least squares estimators for the two hedonic regression models. In particular, we have:

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

Hence, exponentiating both sides of (28) gives us $\exp[c_{ts}] = 1/\exp[c_{ts}]$ and this equation is equivalent to (45) using definitions (20) and (23). Thus we have shown that the estimator of price change P(s,t) defined by (20) (which corresponds to the least squares estimators of the initial log hedonic regression model defined by (9) and (10) with $f(p) \equiv \ln p$) is equal to the reciprocal of the estimator of price change P(t,s) defined by (23) (which corresponds to the second log hedonic regression model defined by (21) and (22) so that the two bilateral dummy variable hedonic regressions satisfy the time reversal test (24).

If it is desired to avoid the use of nonlinear regression techniques, then the results in this section 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.

4. 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 2 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:

$$(29)\ (v_k^s)^{1/2} \ln p_k^s = (v_k^s)^{1/2}[\beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s\ ; \qquad k \in S(s);$$
$$(30)\ (v_k^t)^{1/2} \ln p_k^t = (v_k^t)^{1/2}[\gamma_{st} + \beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t\ ; \qquad k \in S(t);$$

where the model sales values for period t, v_k^t, were defined by (6) 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 (29) and (30) is the bilateral counterpart to our single equation weighted hedonic regression model that was studied in section 2 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 (29) and (30) is unlikely to be satisfied.[21] Hence, it is necessary to pick new weights that will eliminate this problem.

In order to address the above problem, we first define the *period t expenditure share of model k* as follows:

$$(31)\ s_k^t \equiv p_k^t q_k^t / \textstyle\sum_{i\in S(t)} p_i^t q_i^t\ ; \qquad k \in S(t).$$

Our initial 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 as the weights in (29) and (30)

[21] From the viewpoint of the descriptive statistics approach, if we want the weights for each period to be equally important or representative in the regression, then it is natural to require that the weights to sum to one for each period.

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:[22]

$$(32)\ (s_k^s)^{1/2} \ln p_k^s = (s_k^s)^{1/2}[\beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s\ ; \qquad k\in S(s);$$
$$(33)\ (s_k^t)^{1/2} \ln p_k^t = (s_k^t)^{1/2}[\gamma_{st} + \beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t\ ; \qquad k\in S(t);$$

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 (32) and (33), 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:

$$(34)\ P_1(s,t) \equiv \exp[c_{st}].$$

It can be shown that $P_1(s,t)$ defined by (34) in this section has the same desirable property that P(s,t) defined by (20) 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.[23]

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 (34) 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.[24] Hence in place of the model defined by (32) and (33), 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 (32) are replaced by the following two sets of equations:

$$(35)\ (s_k^s)^{1/2} \ln p_k^s = (s_k^s)^{1/2}[\beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s\ ; \qquad k\in[S(s)\sim S(t)];$$
$$(36)\ [(1/2)(s_k^s+s_k^t)]^{1/2} \ln p_k^s = [(1/2)(s_k^s+s_k^t)]^{1/2}\ [\beta_0 + \textstyle\sum_{n=1}^N f_n(z_{kn}^s)\beta_n] + \varepsilon_k^s\ ;\ k\in S(s)\cap S(t).$$

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 (35). 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

[22] Diewert (2005) considered a model similar to (32) and (33) except that all of the explanatory variables were dummy variables and he 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.
[23] See Proposition 1 in the Appendix.
[24] See Diewert (2002; 578-584) and the ILO (2004; 301-304 and 322-324).

included in equations (36). Similarly, the old period s equations (33) are replaced by the following two sets of equations:[25]

(37) $(s_k^t)^{1/2} \ln p_k^t = (s_k^t)^{1/2}[\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t$; $k\in[S(t)\sim S(s)]$;
(38) $[(1/2)(s_k^s+s_k^t)]^{1/2} \ln p_k^t = [(1/2)(s_k^s+s_k^t)]^{1/2} [\gamma_{st} + \beta_0 + \sum_{n=1}^N f_n(z_{kn}^t)\beta_n] + \varepsilon_k^t$; $k\in S(s)\cap S(t)$.

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 (37). 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 (38). 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 (35)-(38), 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:

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

It can be shown that $P_2(s,t)$ defined by (39) 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.[26] Hence *it appears that the weighted hedonic regression model defined by (35)-(38) 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 (35)-(38) becomes a special case of our earlier model defined by (32)-(33).

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:

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

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 (32) and

[25] Note that the "mixed" period s share weights used in (35) and (36) and the "mixed" period t share weights used in (37) and (38) do not necessarily sum to one whereas the period s and t weights used in (32) and (33) respectively did sum to one for each period.
[26] This follows from Corollary 5.3 in the Appendix.

(33) and (35)-(38) 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 axiomatic 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. Thus either index can be justified.

5. The Pure Dummy Variable Adjacent Period Hedonic Regression Model

In this section, we specialize the results in the previous section in order to provide a generalization (to the weighted case) of a model due to Aizcorbe, Corrado and Doms (2000), which introduced a dummy variable for each model in addition to the usual time dummies. Thus their model had no other characteristics other than these model specific dummy variables.[27] We can use the results in the previous section to see how the *weighted* ACD model works in the case of two periods.

We will work with the first share weighted model defined by (32) and (33) in the previous section, except that the N characteristics functions f_n are now assumed to be dummy variables.[28] Thus we assume that there are a total of N different models sold in periods s and t. The old period s equations (32) are broken up into two sets of equations, (41) and (42), where the models k which appear in (41) are present in both periods s and t and the models m which appear in (42) are present in period s and not period t:

(41) $(s_k^s)^{1/2} \ln p_k^s = (s_k^s)^{1/2}\beta_k + \varepsilon_k^s$; $\quad k \in S(s)\cap S(t)$;
(42) $(s_m^s)^{1/2} \ln p_m^s = (s_m^s)^{1/2}\beta_m + \varepsilon_m^s$; $\quad m \in [S(s)\sim S(t)]$.

Similarly, the old period t equations (33) are broken up into two sets of equations, (43) and (44), where the models k which appear in (43) are present in both periods s and t and the models n which appear in (44) are present in period t and not period s:

[27] Their model is equivalent to the Country Product Dummy model used in making international comparisons of prices that was pioneered by Summers (1973). For extensions of this model to the weighted case, see Diewert (2004) (2005).
[28] We also need to set β_0, the constant term in the regression, equal to zero in order to identify all of the parameters in this pure dummy variable hedonic regression.

(43) $(s_k^t)^{1/2} \ln p_k^t = (s_k^t)^{1/2}[\gamma_{st} + \beta_k] + \varepsilon_k^t$; $k \in S(s) \cap S(t)$;
(44) $(s_n^t)^{1/2} \ln p_n^t = (s_n^t)^{1/2}[\gamma_{st} + \beta_n] + \varepsilon_n^t$; $n \in [S(t) \sim S(s)]$.

Now run a least squares regression on the model defined by (41)-(44). Denote the least squares estimates for β_n by b_n for n = 1,....,N and the estimate for γ_{st} by c_{st}. Use the first order necessary (and sufficient) conditions for the least squares minimization problem for the b_n to solve for each b_n in terms of c_{st} and then substitute these expressions for the b_n into the first order condition for the c_{st} parameter. The resulting equation simplifies to:

(45) $c_{st} = W^{-1} \sum_{k \in S(s) \cap S(t)} [s_k^s + s_k^t]^{-1} s_k^s s_k^t \ln[p_k^t/p_k^s]$

where

(46) $W \equiv \sum_{k \in S(s) \cap S(t)} [s_k^s + s_k^t]^{-1} s_k^s s_k^t$.

Thus c_{st}, the log of the hedonic price index going from period s to t, is equal to a weighted average (where the weights are positive and sum to one) of the log price ratios, $\ln[p_k^t/p_k^s]$, over all of the models k that are present in both periods. *Note that this pure dummy variable hedonic model boils down to a (weighted) matched model price index*.

If we drop the share weights in (41)-(44) and simply run an unweighted regression model, then we can use the above algebra by simply setting each "share" equal to 1 and we find that

(47) $c_{st} = (1/M) \ln[p_k^t/p_k^s]$

where M is the number of models that are present in both periods s and t. This captures the original ACD model for the case of only two periods. Thus for the case of only two time periods, the Aizcorbe, Corrado and Doms (2000) model of hedonic price change reduces to a statistical agency matched model estimate of price change; i.e., *their hedonic estimate of the price change going from period s to t is equal to the equally weighted geometric mean of the price relatives of the models that are present in both periods*.

The exact equivalence of the ACD measure of price change in the two period case to the equally weighted geometric mean of the price relatives of the models that are present in both periods does not carry over to the case where there are more than 2 periods. However, it is likely that even in the many period case, the ACD measures of price change will have a tendency to follow the matched model results. In any case, the ACD measures of price change can be quite different from what a hedonic model with continuous characteristics would yield.[29]

6. Conclusion

[29] See Triplett and McDonald (1977), Triplett (2004) and Silver and Heravi (2002) (2003) (2004) (2005) for some empirical evidence and explanation on this point.

We conclude this chapter by noting that we have not resolved all of the issues surrounding the question as to whether hedonic regressions should be weighted according to their economic importance or not. There is a tension between the index number approach to hedonic regressions and the econometric approach. This tension is best described by Triplett (2004), who wrote the most systematic discussion of the weighting issue to date.[30] It is worth quoting part of Triplett's conclusion on the weighting issue:

"Dickens (1990) contends that when weighted and unweighted regression estimates differ, it is a sign of specification error—in our context, an example would be a hedonic function in which crucial characteristics variables were missing. Missing information on software characteristics and some characteristics of hardware are common in hedonic investigations so hedonic function sensitivity to weighting in the presence of missing variables is consistent with Dickens' contention. On Dickens' analysis, in a properly specified hedonic function, weighted and unweighted regressions should not differ." Jack E. Triplett (2004; 193-194).

Thus from the econometrics point of view, the emphasis is on the statistical model: its accurate specification and its most efficient estimation. However, from the viewpoint of the price statistician, the model is not the most important consideration: the most important consideration is to obtain an overall measure of price change over two periods over some domain of definition of admissible prices and transactions involving those prices. Thus the price statistician takes a *descriptive statistics perspective* whereas the econometrician takes a *statistical model and estimation perspective*.[31] The difference between the two approaches can be illustrated in the matched model context. In this context, the econometrician running a hedonic regression model in this context would probably follow Dickens advice and focus on a simple unweighted model of the type defined by (9) and (10) in section 3 and would end up with the equally weighted geometric mean of the price relatives of the matched models (the Jevons index) as the measure of overall price change between the two periods. However, the descriptive statistics price statistician might run the weighted model defined by (35)-(38) in section 4 and would end up with the Törnqvist price index as the measure of overall price change.[32] In many cases, the two estimates of overall price change could be quite different but from the viewpoint of standard index number theory, the Törnqvist index is clearly preferable to the Jevons index.

The theory of hedonic regressions leaves a great deal of leeway open to the empirical investigator with respect to the details of implementation of the models. Our strategy in this chapter has been to use some of the ideas that are present in the test approach to

[30] Silver and Heravi (2004) also discuss alternative approaches to weighting in hedonic regressions in a systematic manner.

[31] Triplett (2004; 190) notes that the econometric approach can be used in a weighted context so the two perspectives *can* be complementary: "However, the dummy variable method is only one method for estimating hedonic price indexes and it is the only one where weights for the index and weights for the hedonic function imply the same questions. Chapter 3 describes three other methods for estimating hedonic price indexes—the characteristics price index method, the hedonic imputation method and the hedonic quality adjustment method—in all of which a weighted index number can be produced using an unweighted hedonic function."

[32] See Diewert (2004; section 8) for a method for converting weighted models like (32)-(33) or (35)-(38) into descriptive statistics models similar to Theil's (1967; 138) so that c_{st}^* can be interpreted as a descriptive statistics measure of overall log pure price change between the two periods.

index number theory in an attempt to work out some of the axiomatic properties of adjacent period time dummy hedonic regression models in an attempt to cast some light on the issue of weighting. Our results are not definitive but perhaps they can cast some light on the consequences of different types of weighting.[33]

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.[34]

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 (32) and (33) in section 4. 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 (32) and the period t share weights s_k^t in (33) 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:

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

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:[35]

[33] It is encouraging to the author that the earlier version of this paper, Diewert (2003b), has been extensively used by Silver and Heravi (2004) (2005) and Haan (2003) and the work of these authors has influenced statistical agency practice in the UK and the Netherlands.

[34] The extensive work by Jack Triplett (2004) and Silver and Heravi (2001) (2002) (2003) (2004) (2005) should help form this consensus.

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

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:

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

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[36] 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 $P(p,p,q,q) = 1$.

(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:

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

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:

---

[35] 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.

[36] 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) von Auer (2001) and the ILO (2004; 292-306).

(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).[37]

[37] In keeping with the test approach in bilateral index number theory, 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.

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:

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

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

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 $c_{st}^{**\prime}$, 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 $c_{st}^{**\prime} = c_{st}^*$, and the c_{st} solution to (A12), which we denote by c_{st}^{**}, satisfies (A14):

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

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).[38]

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

[38] 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.

weights will not be changed by this change in the period s prices and so the new minimization problem will be:

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

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

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^{**}:

$$\text{(A24)}\quad \min_{b\text{'s and } c} \{\textstyle\sum_{k\in S(s)} w_k^s[\ln p_k^s - c_{ts} - b_0 - \sum_{n=1}^N f_n(z_{kn}^s)b_n]^2$$

$$+ \sum_{k\in S(t)} w_k^t[\ln p_k^t - b_0 - \sum_{n=1}^N f_n(z_{kn}^t)b_n]^2\}$$

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

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:

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

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):

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

Using definition (A9), we have:

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

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:[39]

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

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:

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

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

[39] 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 (2004; 51-52). The technique of proof used in this Proposition was used in section 4 of Diewert (2003a).

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:

$$\text{(A31)}\ 0 = \sum_{k\in S(s)} w_k^s \ln p_k^s + \sum_{k\in S(t)} w_k^t \ln p_k^t - [\sum_{k\in S(t)} w_k^t]\, c_{st}^* - [\sum_{k\in S(s)} w_k^s]\, b_0^* - [\sum_{k\in S(t)} w_k^t]\, b_0^* - \sum_{k\in S(s)} w_k^s \sum_{n=1}^N f_n(z_{kn}^s)b_n^* - \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t)b_n^*\ ;$$
$$\text{(A32)}\ 0 = \sum_{k\in S(t)} w_k^t \ln p_k^t - \sum_{k\in S(t)} w_k^t c_{st}^* - \sum_{k\in S(t)} w_k^t b_0^* - \sum_{k\in S(t)} w_k^t \sum_{n=1}^N f_n(z_{kn}^t)b_n^*.$$

Equation (A32) can be rewritten as:

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

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

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

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

Corollary 5.1: The solution coefficients c_{st}^*, b_0^*, b_1^*,..., b_N^* to (A2) satisfy the following two equations:

$$\text{(A35)}\ c_{st}^* = \{\sum_{k\in S(t)} w_k^t [\ln p_k^t - b_0^* - \sum_{n=1}^N f_n(z_{kn}^t)b_n^*]\}/\sum_{i\in S(t)} w_i^t\ ;$$
$$\text{(A36)}\quad 0 = \{\sum_{k\in S(s)} w_k^s [\ln p_k^s - b_0^* - \sum_{n=1}^N f_n(z_{kn}^s)b_n^*]\}/\sum_{i\in S(s)} w_i^s\ .$$

Proof: (A35) is a rearrangement of (A33) and (A36) is a rearrangement of (A34). Q.E.D.

Corollary 5.2: 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:

$$\text{(A37)}\ c_{st}^* = \sum_{k=1}^K w_k \ln [p_k^t/p_k^s]/[\sum_{j=1}^K w_j]$$

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

Corollary 5.3: 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 (A37) 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. Using Corollary 5.1, we can establish the following formula for c_{st}^*, where the constant term b_0^* has been eliminated from (A29):

$$\text{(A38)}\quad c_{st}^* = \sum_{k\in S(t)} s_k^t [\ln p_k^t - \sum_{n=1}^N f_n(z_{kn}^t)b_n^*] - \sum_{k\in S(s)} s_k^s [\ln p_k^s - \sum_{n=1}^N f_n(z_{kn}^s)b_n^*]$$

where $s_k^t \equiv w_k^t/^*\sum_{i\in S(t)} w_i^t$ for $k\in S(t)$ and $s_k^s \equiv w_k^s/^*\sum_{i\in S(s)} w_i^s$ for $k\in S(s)$ so that the weights in (A38) now sum up to one in each period. Thus the two weighted sums in (A38) become weighted *averages* of the logarithms of quality adjusted prices. If the characteristics transformation functions f_n are log functions, then the interpretation of (A38) becomes particularly transparent.[40]

References

Aizcorbe, A., C. Corrado and M. Doms (2000), "Constructing Price and Quantity Indexes for High Technology Goods", Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System, Washington DC.

Allen, R.C. and W.E. Diewert (1981), "Direct versus Implicit Superlative Index Number Formulae", *The Review of Economics and Statistics* 63, 430-435.

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.

Clements, K.W. and H.Y. Izan (1981), "A Note on Estimating Divisia Index Numbers", *International Economic Review* 22, 745-747.

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

[40] This type of decomposition can be traced back to Court (1939; 108) as his hedonic suggestion number one. The same decomposition was suggested by 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).

Dickens, W.T. (1990), "Error Components in Grouped Data: Is it Ever Worth Weighting?", *Review of Economics and Statistics* 72, 328-333.

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. (1995), "On the Stochastic Approach To Index Numbers", Discussion Paper 95-31, Department of Economics, University of British Columbia, Vancouver, Canada. http://www.econ.ubc.ca/diewert/9531.pdf

Diewert, W.E. (2002), "Harmonized Indexes of Consumer Prices: Their Conceptual Foundations", *Swiss Journal of Economics and Statistics* 138:4, 547-637. http://www.econ.ubc.ca/diewert/other.htm

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

Diewert, W.E. (2003b), "Hedonic Regressions: A Review of Some Unresolved Issues", Paper presented at the Seventh Meeting of the International Working Group on Price Indices, (Ottawa Group), 27-29 May 2003, INSEE, Paris. http://www.insee.fr/en/nom_def_met/colloques/ottawa/ottawa_papers.htm

Diewert, W.E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Discussion Paper No. 04-16, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E. (2005), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", *Review of Income and Wealth*, forthcoming.

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.

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

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Geneva: International Labour Office.

Koskimäki, T. and Y. Vartia (2001), "Beyond Matched Pairs and Griliches Type Hedonic Methods for Controlling Quality Changes in CPI Sub-Indices", paper presented at the 6th Ottawa Group Meeting, Canberra, Australia, April 2-6, 2001.

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.

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

Selvanathan, E. A. and D. S. Prasada Rao (1994), *Index Numbers: A Stochastic Approach*, Ann Arbor: The University of Michigan Press.

Silver, M. and Heravi, S. (2002), "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: European Central Bank.

Silver, M. and Heravi, S. (2003), "The Measurement of Quality-Adjusted Price Changes", pp. 277-316 in *Scanner Data and Price Indexes*, Mathew Shapiro and Rob Feenstra (eds.), National Bureau of Economic Research, Studies in Income and Wealth, vol. 61, Chicago: University of Chicago Press.

Silver, M. and Heravi, S. (2004), “The Difference Between Hedonic Imputation Indexes and Time Dummy Hedonic Indexes”, paper presented at the NBER/CRIW Conference *Price Index Concepts and Measurement*, Vancouver, June 28-29, 2004.

Silver, M. and Heravi, S. (2005), “A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment using Scanner Data”, *Journal of Business and Economic Statistics* 23:3, 269-281.

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

Summers, R. (1973), “International Comparisons with Incomplete Data”, *Review of Income and Wealth* 29:1, 1-16.

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

Triplett, J. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, Directorate for Science, Technology and Industry Working Paper 2004/9, Paris: OECD. http://www.oecd.org.dataoecd/37/31/33789552.pdf

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.