Quality Adjustment and Hedonics: A Unified Approach

Erwin Diewert,[1]
Discussion Paper 19-01,
Vancouver School of Economics,
University of British Columbia,
Vancouver, B.C., Canada, V6T 1L4.

January 29, 2019.

Abstract

The paper takes a consumer demand perspective to the problem of adjusting product prices for quality change. The various approaches to the problem of quality adjustment can be seen as special cases of the general framework. The special cases include the use of inflation adjusted carry forward and carry backward prices, the use of hedonic regressions and the estimation of Hicksian reservation prices.

Keywords

Quality adjustment, hedonic regressions, reservation prices, consumer theory, time product dummy regressions, scanner data.

JEL Classification Numbers

C43, C81, E31.

[1] W. Erwin Diewert: Vancouver School of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and the School of Economics, UNSW Sydney, NSW 2052, Australia (erwin.diewert@ubc.ca). The author thanks Jan de Haan, Adam Gorajek and Chihiro Shimizu for helpful comments, Jerry Hausman, and Mick Silver for helpful discussions over the years and he gratefully acknowledges the financial support of a SSHRC grant.

Table of Contents

1. Introduction

This paper will attempt to place most methods used by statistical agencies to quality adjust prices into a common economic framework. The economic framework is based on purchasers maximizing a linearly homogeneous utility function subject to a budget constraint on their purchases of a group of related products. This framework is far from a perfect description of reality but it captures an important empirical phenomenon: when the price of a product drops a lot, purchasers of the product buy more of it! Moreover, the theory allows us to provide a welfare interpretation for the quantity indexes which are generated by this approach.

The theory of quality adjustment to be presented in this paper is meant to be applied at the level where subindexes are constructed at the first stage of aggregation; i.e., at what is called the *elementary level of aggregation* by price statisticians. Furthermore, the methods for quality adjustment to be discussed in this paper are largely aimed at the *scanner data context*; i.e., we will assume that the statistical agency has access to detailed price and quantity (or value) information at the product code level, either from retail outlets or from the detailed purchases of a group of similar households.[2] Thus our focus will be on both the construction of consumer price indexes at the elementary level as well as on the companion consumer quantity indexes.

The assumption of linearly homogeneous utility or valuation functions is an important restriction so one may ask why impose it? The reason is that economic models constructed by private and public sector economists generally do not make use of disaggregated information; instead, they use the elementary indexes that are produced by national statistical agencies in their models. However, the price levels that correspond to these elementary indexes are treated as "normal" prices by applied economists; i.e., the elementary prices are not regarded as prices that are conditional on particular levels of the corresponding quantity levels. In order to construct unconditional price levels, we need to assume that the underlying aggregator or utility functions are linearly homogeneous.[3]

Marshall (1887) was one of the first to introduce *the new goods problem*: how exactly should price indexes be adjusted to account for the introduction of new and hopefully improved products? [4] Marshall suggested that chaining period to period indexes would

---

[2] As cash transactions diminish in importance, credit and debit card companies will have detailed price and quantity information on household purchases. Once this information on consumer transactions also includes product bar codes, statistical agencies will eventually be able to access this information and use it to produce high quality consumer price indexes.

[3] The underlying index number theory using linearly homogeneous aggregator functions was developed by Shephard (1953), Samuelson and Swamy (1974) and Diewert (1976). This theory will be explained in section 2 below.

[4] "This brings us to consider the great problem of how to modify our unit so as to allow for the invention of new commodities. The difficulty is insuperable, if we compare two distant periods without access to the detailed statistics of intermediate times, but it can be got over fairly well by systematic statistics." Alfred Marshall (1887; 373). Lehr (1885; 45-46) also introduced the chain system as a way of mitigating the new goods problem. For more on the early history of the new goods problem, see Diewert (1993; 59-63).

provide a partial solution to the problem. Keynes (1909) endorsed Marshall's suggestion as a step in the right direction but noted that chaining alone will not solve the fundamental problem: increased product choice will generally increase the utility of purchasers of products but it is very difficult to measure this increase.[5] This is the essence of the *quality adjustment problem*; how can statistical agencies construct price and quantity indexes over two or more periods when there are new and disappearing products?

Hicks (1940; 114) suggested a general approach to this measurement problem in the context of the economic approach to index number theory. His approach was to apply normal index number theory but estimate (somehow) hypothetical prices that would induce utility maximizing purchasers of a related group of products to demand 0 units of unavailable products.[6] With these *reservation* or *imputed prices* in hand, one can just apply normal index number theory using the augmented price data and the observed quantity data (which impute zero quantities to unavailable products). This is the economic framework we will use in this paper.[7] The practical problem facing statistical agencies is: *how exactly are these reservation prices to be estimated*?

The approach to the estimation of reservation prices that will be taken below is to use consumer demand theory to estimate preferences. Suppose that purchasers maximize a utility function $Q(q)$ subject to the budget constraint $p\cdot q \equiv \Sigma_{n=1}^{N} p_nq_n = v > 0$ where the price and quantity of commodity n are p_n and q_n for $n = 1,...,N$. Define the price and quantity vectors $p \equiv [p_1,...,p_N]$ and $q \equiv [q_1,...,q_N]$. Suppose that p, q and v are observed and q is a solution to the utility maximization problem $\max_q \{Q(q) : p\cdot q = v\}$. Then given a functional form for Q, the solution to the utility maximization problem will satisfy the usual consumer demand functions, $q_n = f_n(p,v)$ for $n = 1,...,N$ where $f_n(p,v)$ is the nth consumer demand function. Given price and quantity for many periods, the unknown parameters for the utility function that are imbedded in these consumer demand functions can be estimated using econometric methods. Duality theory can be used to

---

[5] "The [chaining] method has another advantage. It enables us to introduce new commodities and to drop others which have fallen out of use. ... For most practical purposes, therefore, this is the method to be recommended. ... Yet we must not exaggerate its merits." John M. Keynes (1909; 80). "We cannot hope to find a ratio of equivalent substitution for gladiators against cinemas, or for the conveniences of being able to buy motor cars against the conveniences of being able to buy slaves." John M. Keynes (1930; 96).

[6] "The same kind of device can be used in another difficult case, that in which new sorts of goods are introduced in the interval between the two situations we are comparing. If certain goods are available in the II situation which were not available in the I situation, the p_1's corresponding to these goods become indeterminate. The p_2's and q_2's are given by the data and the q_1"s are zero. Nevertheless, although the p_1's cannot be determined from the data, since the goods are not sold in the I situation, it is apparent from the preceding argument what p_1's ought to be introduced in order to make the index-number tests hold. They are those prices which, in the I situation, would *just* make the demands for these commodities (from the whole community) equal to zero." John R. Hicks (1940; 114). von Hofsten (1952; 95-97) extended Hicks' methodology to cover the case of disappearing goods as well.

[7] Two major problems with this framework are: (i) it does not take into account the fact that purchasers may stockpile goods on sale and this will affect demand in subsequent periods and (ii) the introduction of a new revolutionary product may change purchaser preferences over existing goods. However, until a better welfare oriented model of purchaser behavior comes along, we are stuck with using the Hicksian approach.

simplify the derivation of the consumer demand functions.[8] This is the approach used by Hausman (1981) (1996) (1999) (2003) used to estimate reservation prices. However, the econometrics of this method are complex. To illustrate these problems, suppose in the first period of the sample period, product 1 was not available. The observed demand for product 1 in period 1 is zero. Thus the first estimating equation in the sample would take the form $0 = f_1(p_1^{1*},p_2^1,...,p_N^1,v^1) + e_1^1$ where $f_1(p,v)$ is the demand function for commodity 1, $p_2^1,...,p_N^1$ are the observed prices for products 2,3,...,N in period 1, v^1 is the observed period 1 expenditure on the N products, e_1^1 is an error term and p_1^{1*} is the unknown period 1 reservation price for product 1. It can be seen that p_1^{1*} is now an extra parameter that must be estimated. Hence the usual approach that conditions on prices (on the right hand sides of the estimating equations) and treats quantities as random variables on the left hand sides of the estimating equations does not apply due to the endogeneity of the reservation price. Moreover, the variable on the left hand side of the above equation is 0 and this is not a random variable. Thus simple econometric techniques cannot be used in this situation.

To deal with the above econometric problem, one can abandon the estimation of traditional consumer demand functions and switch to the estimation of the system of *inverse consumer demand functions*. The nth inverse demand function gives the observed price for product n, p_n, as a function of the vector of quantities chosen by the purchasers, q and total expenditure on the products v; i.e., we have $p_n = g_n(q,v)$ for n = 1,...,N where g_n is the nth inverse demand function.[9] Again suppose product 1 was not available in period 1. Then the first inverse demand function in period 1 becomes $p_1^{1*} = g_1(0,q_2^1,...,q_N^1,v^1) + e_1^1$ using the notation in the previous paragraph. Thus we simply drop this equation from the system of inverse demand estimating equations and use the remaining equations to estimate the unknown parameters in the direct utility function Q(q). Once these unknown parameters have been estimated, the period 1 reservation price for product 1 can be defined as $p_1^{1*} = g_1(0,q_2^1,...,q_N^1,v^1)$. This methodology will be described in section 14 in more detail.[10]

It turns out that a special case of this inverse demand function methodology is the case of a *linear utility function*; i.e., set $Q(q) = \Sigma_{n=1}^N \alpha_n q_n \equiv \alpha \cdot q$ where the α_n are *quality adjustment factors*. Thus α_n gives the increase in utility of purchasers due to the acquisition of an extra unit of product n. The case of a linear utility function will be used as an underlying economic model in sections 3 and 5-12. It turns out that the assumption of a linear utility function allows us to give an economic interpretation to most hedonic regression models.

In sections 3 and 4, we apply the linear utility function assumption to some special situations where it is possible to generate missing prices without using any econometrics.

[8] See for example Diewert (1974; 120-133).

[9] Suppose that the utility function Q(q) is differentiable and linearly homogeneous and we have an interior solution to the purchaser's utility maximization problem. Then using Wold's (1944; 69-71) identity, $p_n = [\partial Q(q)/\partial q_n]v/Q(q) \equiv g_n(q,v)$ for n = 1,...,N. We will derive these equation in more detail in section 2 below.

[10] This methodology was first suggested by Diewert (1980; 498-503) and implemented by Diewert and Feenstra (2017).

These sections introduce inflation adjusted carry forward and carry backward prices which have been used for many years by statistical agencies to replace missing prices.[11]

In section 5, we also assume an underlying linear utility function but we no longer assume that the underlying economic model holds exactly. Thus error terms make their appearance in this section (and in subsequent sections). The resulting model is the *time product dummy hedonic regression model*. This model is an application of Summer's (1973) *country product dummy model* to the time series context. The underlying time product dummy hedonic regression model is $p_{tn} = \pi_t\alpha_n$ for n = 1,...,N and t = 1,...,T where the α_n are the *quality adjustment factors* that appear in the purchasers' linear utility function and the π_t turn out to be period t *aggregate price levels*.[12] However, in real life applications, these equations will not hold exactly and thus it is necessary to introduce error terms. The above exact equations are replaced by $p_{tn} = \pi_t\alpha_n + e_{tn}$ for n = 1,...,N and t = 1,...,T where the e_{tn} are error terms. Estimators for the π_t and α_n are found by minimizing the sum of squared errors, $\Sigma_{t=1}^T \Sigma_{n=1}^N [e_{tn}]^2 = \Sigma_{t=1}^T \Sigma_{n=1}^N [p_{tn} - \pi_t\alpha_n]^2$. However, the resulting estimated price levels, π_t^* for t = 1,...,T, turn out to have unsatisfactory properties from the viewpoint of the test or axiomatic approach to index number theory. Thus we take logarithms of both sides of the exact equations and add error terms to the resulting equations. New estimates for the π_t are obtained by minimizing the new sum of squared errors, $\Sigma_{t=1}^T \Sigma_{n=1}^N [\ln p_{tn} - \rho_t - \beta_n]^2$, where $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and $\beta_n \equiv \alpha_n$ for n = 1,...,N. The resulting price level estimates π_t^* have much better axiomatic properties; they turn out to be the geometric means of the quality adjusted prices for period t; i.e., $\pi_t^* = \Pi_{n=1}^N [p_{tn}/\alpha_n^*]^{1/N}$ for t = 1,...,T.[13]

Note that we started with a model based on the economic approach to index numbers but after recognizing that the economic model was unlikely to hold exactly, we introduced error terms and hence we drifted into the stochastic approach to index numbers. Finally, the test or axiomatic approach to index number theory was applied to decide on how to transform the exact equations into a form that leads to more satisfactory estimated price levels. Thus the overall approach includes elements of all three major approaches to index number theory.[14]

---

[11] See von Hofsten (1952), Triplett (2004), de Haan and Krsinich (2012; 31-32) and Diewert, Fox and Schreyer (2017).

[12] A *bilateral price index* between period t relative to period r is defined as the ratio of the relevant price levels, π_t/π_r.

[13] See equations (36) below.

[14] It should also be noted that our stochastic approach is not very sophisticated from an econometric perspective; i.e., we simply use a least squares approach to generate estimated price levels without considering the possibility of bias in our estimates. Thus our stochastic approach is more of a descriptive statistics approach. Our approach to bias and efficiency of estimators is similar to that of de Haan and Krsinich (2018; 770): "Our preference for the index number perspective can have implications for the efficiency of the estimators. It may be that our choice of weights introduces or amplifies heteroskedasticity and leads to estimators that are less efficient than strictly necessary. But a loss of efficiency is not as bad as ending up with indexes that are not grounded in index number theory." For more rigorous econometric approaches, see Rao and Hajargasht (2016) and Gorajek (2018).

Section 6 uses the same model as in section 5 except that the analysis in section 6 allows for missing observations.

The time product dummy regression models studied in sections 5 and 6 suffer from a major defect: if prices are the same in periods r and t, then the bilateral price index, π_t^*/π_r^* generated by the models in these sections boils down to the Jevons (1863) (1865) index, which is the unweighted[15] geometric mean of the price ratios p_{tn}/p_{rn} for the products that are present in the two periods. This index is a very good one if quantity or value information is not available, but it is not satisfactory when such information is available. Thus in subsequent sections, we study hedonic regression models that use weights.

In section 7, we recognize that there are three sets of weights that could be used to recognize the *economic importance* of prices in the weighted least squared minimization problems that are the counterparts to the unweighted least squared minimization problems in sections 5. The three sets of possible weights for p_{tn} are: (i) the corresponding quantity sold, q_{tn}; (ii) the corresponding value of sales of the product, v_{tn} or (iii) the corresponding sales share for product n in period t, s_{tn}. Thus the general class of weighted least squared minimization problems studied in section 7 that are counterparts to the equally weighted problems in section 5 are of the form $\Sigma_{t=1}^{T} \Sigma_{n=1}^{N} w_{tn}[e_{tn}]^2 = \Sigma_{t=1}^{T} \Sigma_{n=1}^{N} w_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2$ where w_{tn} is equal to q_{tn}, v_{tn} or s_{tn}. In section 7, we restrict attention to the case of only 2 periods so that T = 2. In this bilateral index number context, we consider a fourth set of weights where $w_{1n} = w_{2n} \equiv ½(s_{1n} + s_{2n})$ for n = 1,...,N. This last choice of weights gives the best result. This section shows how the strategic selection of weights can transform a rather poor price index π_2^*/π_1^* into a superlative price index.[16] The results in this section illustrate the importance of weighting.

Section 8 generalizes the section 7 bilateral model to the case where there are missing observations. The results in this section are rather complicated. The bottom line is that the bilateral time product dummy hedonic regression model with missing observations is likely to be dominated by the use of the maximum overlap Fisher (1922) index methodology that is explained in section 4.

Section 9 considers the weighted time product dummy model for T periods with missing observations. The resulting price levels have satisfactory axiomatic properties but this method has an important drawback: a product that is available only in one period out of the T periods has no influence on the estimated aggregate price levels π_t^* for all periods. Thus the introduction of a new product in period T will have no effect on the estimated price level for period T, π_T^*. This goes against the spirit of the Hicksian approach to the treatment of new goods. The hedonic regression models considered in the following two sections do not suffer from this drawback.

Sections 10 and 11 deal with hedonic regression models that make use of information on the characteristics of the N products under consideration. The models in these two

[15] When we speak of unweighted means, we mean equally weighted means.
[16] See Diewert (1976) for the definition of a superlative price index.

sections are more satisfactory than the weighted time product dummy model discussed in section 9 because now isolated prices play a role in the determination of the estimated price levels π_t^* for t = 1,...,T. However the hedonic regression models considered in sections 10 and 11 do require information on product characteristics, information which may be difficult to collect. The important results obtained by de Haan and Krsinich (2018) using this class of hedonic regression models applied to electronic products are discussed in section 11. They compare weighted and unweighted versions of the same hedonic regression models and show that weighting leads to improved results.

The problems raised by taste change in the case where T = 2 are addressed in section 12. The treatment of the problem in this section is due to Diewert, Heravi and Silver (2009) and it uses the tastes of each period to construct a bilateral price index between the two periods. The two indexes are then averaged to form a final index.

Finally, in sections 13 and 14, two alternative methods for constructing reservation prices in situations are discussed. In these models, the underlying utility function is *not* assumed to be a linear function. In section 13, the reservation price model due to Feenstra (1994) is presented. This model assumes that the underlying preferences are CES (Constant Elasticity of Substitution).[17] The model presented in section 14 assumes that the underlying preferences are a certain flexible functional form (that is exact for the Fisher (1922) ideal quantity index) and this model is due to Diewert and Feenstra (2017).

Section 15 offers a few conclusions.

2. A Framework for Evaluating Quality Change In the Scanner Data Context

In this section, we attempt to provide a framework for the construction of consumer price and quantity indexes in the scanner data context using the economic approach to index number theory. We assume that transactions data for the sales or purchases of N products over T time periods are available.[18] The N products will typically be a group of related products so that the goal is the construction of price and quantity indexes at the first stage of aggregation. The transactions data are aggregated over time within each period. Let $p^t \equiv [p_{t1},...,p_{tN}]$ and $q^t \equiv [q_{t1},...,q_{tN}]$ denote the price and quantity vectors for time periods t = 1,...,T. The period t quantity for product n, q_{tn}, is equal to total purchases of product n by purchasers or to the sales of product n by the outlet (or group of outlets) for period t, while the period t price for product n, p_{tn}, is equal to the value of sales (or purchases) of product n in period t, v_{tn}, divided by the corresponding total quantity sold (or purchased), q_{tn}. Thus $p_{tn} \equiv v_{tn}/q_{tn}$ is the *unit value price* for product n in period t for t = 1,...,T and n = 1,...,N. In this section, we assume that all prices, quantities and values are positive; in subsequent sections, this assumption will be relaxed.

[17] See Arrow, Chenery, Minhas and Solow (1961) for the first use of this functional form in the economics literature.

[18] The data could be price and quantity (or value and quantity) on sales of the N products from a retail outlet (or group of outlets in the same region) or it could be price and quantity data for the purchases of the N products by a group of similar households.

Let $q \equiv [q_1,...,q_N]$ be a generic quantity vector. In order to compare various methods for comparing the value of alternative combinations of the N products, it is necessary that a *valuation function* or *aggregator function*, Q(q), exist. This function allows us to value alternative combinations of products; if $Q(q^2) > Q(q^1)$, then purchasers of the products place a higher utility value on the vector of purchases q^2 than they place on the vector of purchases q^1. The function Q(q) can also act as an *aggregate quantity level* for the vector of purchases, q. Thus $Q(q^t)$ can be interpreted as an aggregate quantity level for the period t vector of purchases, q^t, and the ratios, $Q(q^t)/Q(q^1)$, t = 1,..,T, can be interpreted as *fixed base quantity indexes* covering periods 1 to T.

In the following analysis, we assume that Q(q) has the following properties: (i) Q(q) > 0 if $q >> 0_N$;[19] (ii) Q(q) is nondecreasing in its components; (iii) $Q(\lambda q) = \lambda Q(q)$ for $q \geq 0_N$ and $\lambda \geq 0$; (iv) Q(q) is a continuous concave function over the nonnegative orthant. Assumption (iii), linear homogeneity of Q(q), is a somewhat restrictive assumption. However, this assumption is required to ensure that the aggregate price level, P(p,q), that corresponds to Q(q) does not depend on the scale of q.[20] Property (iv) will ensure that the first order necessary conditions for the budget constrained maximization of Q(q) are also sufficient.

Let $p \equiv [p_1,...,p_N] > 0_N$ and $q \equiv [q_1,...,q_N] > 0_N$ with $p \cdot q \equiv \Sigma_{n=1}^N p_n q_n > 0$. Then the *aggregate price level*, P(p,q) that corresponds to the aggregate quantity level Q(q) is defined as follows:

(1) $P(p,q) \equiv p \cdot q/Q(q)$.

Thus the implicit price level that is generated by the generic price and quantity vectors, p and q, is equal to the value of purchases, $p \cdot q$, deflated by the aggregate quantity level, Q(q). Note that using these definitions, the product of the aggregate price and quantity levels equals the value of purchases during the period, $p \cdot q$.

Once the functional form for the aggregator function Q(q) is known, then the *aggregate quantity level for period t*, Q^t, can be calculated in the obvious manner:

(2) $Q^t \equiv Q(q^t)$; t = 1,...,T.

Using definition (1), the corresponding period t aggregate price level, P^t, can be calculated as follows:

(3) $P^t \equiv p^t \cdot q^t/Q(q^t)$; t = 1,...,T.

[19] Notation: $q >> 0_N$ means each component of q is positive, $q \geq 0_N$ means each component of q is nonnegative and $q > 0_N$ means $q \geq 0_N$ but $q \neq 0_N$,

[20] $P(p,q) \equiv p \cdot q/Q(q)$ where $p \cdot q \equiv \Sigma_{n=1}^N p_n q_n$. Thus using property (iii) of Q(q), we have $P(p,\lambda q) = p \cdot \lambda q/Q(\lambda q) = P(p,q)$.

Note that if Q(q) turns out to be a *linear aggregator function*, so that $Q(q^t) \equiv \alpha \cdot q^t = \Sigma_{n=1}^N \alpha_n q_{tn}$, then the corresponding period t price level P^t is equal to $p^t \cdot q^t/\alpha \cdot q^t$, which is a *quality adjusted unit value price level*.

In order to make further progress, it is necessary to make some additional assumptions. The two additional assumptions are: (v) Q(q) is once differentiable with respect to the components of q and (vi) the observed strictly positive quantity vector for period t, $q^t >> 0_N$, is a solution to the following period t constrained maximization problem defined by:[21]

$$(4)\ \max_q \{Q(q) : p^t \cdot q = v^t ; q \geq 0_N\}; \qquad t = 1,...,T.$$

The first order conditions for solving (4) for period t are the following conditions:[22]

$$(5)\ \nabla_q Q(q^t) = \lambda_t p^t ; \qquad t = 1,...,T;$$
$$(6)\ \quad p^t \cdot q^t = v^t ; \qquad t = 1,...,T.$$

Since Q(q) is assumed to be linearly homogeneous with respect to q, Euler's Theorem on homogeneous functions implies that the following equations hold:

$$(7)\ q^t \cdot \nabla_q Q(q^t) = Q(q^t) ; \qquad t = 1,...,T.$$

Take the inner product of both sides of equations (5) with q^t and use the resulting equations along with equations (7) to solve for the Lagrange multipliers, λ_t:

$$(8)\ \lambda_t = Q(q^t)/p^t \cdot q^t \qquad t = 1,...,T$$
$$= 1/P^t \qquad \text{using definitions (3).}$$

Thus if we assume utility maximizing behavior on the part of purchasers of the N products using the collective utility function Q(q) that satisfies the above regularity conditions, then the period t quantity aggregate is $Q^t \equiv Q(q^t)$ and the companion period t price level defined as $P^t \equiv p^t \cdot q^t/Q^t$ is equal to $1/\lambda_t$ where λ_t is the Lagrange multiplier for problem t in the constrained utility maximization problems (4) and where q^t and λ_t solve equations (5) and (6) for period t. Equations (8) also imply that the product of P^t and Q^t is exactly equal to observed period t expenditure v_t; i.e., we have

$$(9)\ P^t Q^t = p^t \cdot q^t = v_t ; \qquad t = 1,...,T.$$

[21] The theory that follows dates back to Konüs and Byushgens (1926). This approach blends standard consumer demand theory based on the maximization of a linearly homogeneous utility function with index number theory. It was further developed by Shephard (1953) (in the context of a producer cost minimization framework) and by Samuelson and Swamy (1974) and Diewert (1976) in the consumer context. The price indexes which result from this theory are special cases of the Konüs (1924) true cost of living index. What is new in the present paper is the application of this theory to hedonic regression models.

[22] Using the assumption of concavity of Q(q) and the assumption that $q^t >> 0_N$, these conditions are also sufficient to solve (4). Notation: $\nabla_q Q(q) \equiv [\partial Q(q)/\partial q_1,..., \partial Q(q)/\partial q_N]$.

Substitute equations (8) into equations (5) and after a bit of rearrangement, the following *fundamental equations* are obtained:[23]

$$(10)\ p^t = P^t\nabla_q Q(q^t)\ ; \qquad t = 1,...,T.$$

In the following section, we will assume that the aggregator function, Q(q) is a linear function and we will show how this assumption along with equations (9) for the case where T = 2 and N = 3 can lead to a simple well known method for quality adjustment that does not involve any econometric estimation of the parameters of the linear function. In subsequent sections, equations (10) will be utilized in the hedonic regression context and finally, in the final sections of the paper, the assumption that Q(q) is a linear function will be relaxed.

3. A Nonstochastic Method for Quality Adjustment: A Simple Model

A major problem that arises when statistical agencies use scanner data to construct an elementary index is that some products are sold or purchased in one period but not in a subsequent period. Conversely, new products appear in the present period which were not present in previous periods. How should price and quantity indexes be constructed under these circumstances? Equations (10) in the previous section can be used to provide an answer to this question.

Consider the special case where the number of periods T is equal to 2 and the number of products in scope for the elementary index is N equal to 3. Product 1 is present in both periods, product 2 is present in period 1 but not in period 2 (a disappearing product) and product 3 is not present in period 1 but is present in period 3 (a new product).[24] We assume that purchasers of the three products behave as if they collectively maximized the following linear aggregator function:

$$(11)\ Q(q_1,q_2,q_3) \equiv \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3$$

where the α_n are positive constants. Under these assumptions, equations (10) written out in scalar form become the following equations:[25]

---

[23] Multiply the right hand side of equation t in (10) by $1 = Q^t/Q(q^t)$ and use $P^tQ^t = v_t$ to obtain the following system of equations: $p^t = v_t\nabla_q Q(q^t)/Q(q^t)$ for t = 1,...,T. For each t, this system of equations is the consumer's system of *inverse demand functions*, that give the period t prices that rationalize the observed period t demands q^t as functions of q^t and period t expenditure v_t. Konüs and Byushgens (1926) obtained a system of equations that are equivalent to this system of inverse demand functions. Linear homogeneity of the utility function is required to obtain these equations and the equivalent equations defined by (9) and (10).

[24] The "new" product may not be a truly new product; it may be the case that product 3 was temporarily not available in period 1. Similarly, product 2 may not permanently disappear in period 2; it may reappear in a subsequent period.

[25] This is a special case of the Time Product Dummy regression model which will be studied in more detail in subsequent sections. Thus equations (12), which are the inverse consumer demand functions that result from the maximization of a linear utility function, lead directly to a particular hedonic regression model. It

(12) $p_{tn} = P^t \alpha_n$; $n = 1,2,3; t = 1,2.$

Equations (12) are 6 equations in the 5 parameters P^1 and P^2 (which can be interpreted as *aggregate price levels* for periods 1 and 2) and α_1, α_2 and α_3, which can be interpreted as *quality adjustment factors* for the 3 products; i.e., each α_n measures the relative usefulness of an additional unit of product n to purchasers of the 3 products. However, product 3 is not observed in the marketplace during period 1 and product 2 is not observed in the marketplace in period 2 and so there are only 4 equations in (12) to determine 5 parameters. However, the P^t and the α_n cannot all be identified using observable data; i.e., if P^1, P^2, α_1, α_2 and α_3 satisfy equations (12) and λ is any positive number, then λP^1, λP^2, $\lambda^{-1}\alpha_1$, $\lambda^{-1}\alpha_2$ and $\lambda^{-1}\alpha_3$ will also satisfy equations (12). Thus it is necessary to place a normalization (like $P^1 = 1$ or $\alpha_1 = 1$) on the 5 parameters which appear in equations (12) in order to obtain a unique solution. In the index number context, it is natural to set the price level for period 1 equal to unity and so we impose the following normalization on the 5 unknown parameters which appear in equations (12):

(13) $P^1 = 1.$

The 4 equations in (12) which involve observed prices and the single equation (13) are 5 equations in 5 unknowns. The unique solution to these equations is:

(14) $P^1 = 1;\ P^2 = p_{21}/p_{11};\ \alpha_1 = p_{11};\ \alpha_2 = p_{12};\ \alpha_3 = p_{23}/(p_{21}/p_{11}) = p_{23}/P^2.$

Note that the resulting *price index*, P^2/P^1, is equal to p_{21}/p_{11}, the price ratio for the commodity that is present in both periods. Thus the price index for this very simple model turns out to be a *maximum overlap price index*.[26]

Once the P^t and α_n have been determined, equations (12) for the missing products can be used to define the following *imputed prices* p_{tn}^* for commodity 3 in period 1 and product 2 in period 2:

(15) $p_{13}^* \equiv P^1\alpha_3 = p_{23}/(P^2/P^1)$; $p_{22}^* \equiv P^2\alpha_2 = (p_{21}/p_{11})p_{12} = (P^2/P^1)p_{12}.$

These imputed prices can be interpreted as Hicksian (1940; 12) *reservation prices*;[27] i.e., they are the lowest possible prices that would deter purchasers from purchasing the products during periods if the unavailable products hypothetically became available.[28]

---

is this result which allows us to claim that our present approach is a way of reconciling hedonic regression models with classical consumer demand theory.

[26] Keynes (1930; 94) was an early author that advocated this method for dealing with new goods by restricting attention to the goods that were present in both periods being compared. He called his suggested method the *highest common factor method*. Marshall (1887; 373) implicitly endorsed this method. Triplett (2004; 18) called it the *overlapping link method*.

[27] Hicks (1940 dealt only with the case of new goods; von Hofsten (1952; 95-97) extended his approach to cover the case of disappearing goods as well.

Note that $p_{13}^{*} = p_{23}/(P^2/P^1)$ is an *inflation adjusted carry backward price*; i.e., the observed price for product 3 in period 2, p_{23}, is divided by the maximum overlap price index P^2/P^1 in order to obtain a "reasonable" valuation for a unit of product 3 in period 1. Similarly, $p_{22}^{*} = (P^2/P^1)p_{12}$ is an *inflation adjusted carry forward price* for product 2 in period 2; i.e., the observed price for product 2 in period 1, p_{12}, is multiplied by the maximum overlap price index P^2/P^1 in order to obtain a "reasonable" valuation for a unit of product 2 in period 2.[29]

Note that the above algebra can be implemented without a knowledge of quantities sold or purchased. Assuming that quantity information is available, we now consider how companion quantity levels, Q^1 and Q^2, for the price levels, P^1 and P^2, can be determined. Note that $q_{13} = 0$ and $q_{22} = 0$ since consumers cannot purchase products that are not available. Use the imputed prices defined by (15) to obtain complete price vectors for each period; i.e., define the period 1 complete price vector by $p^1 \equiv [p_{11}, p_{12}, p_{13}^{*}]$ and the complete period 2 price vector by $p^2 \equiv [p_{21}, p_{22}^{*}, p_{23}]$. The corresponding complete quantity vectors are by $q^1 \equiv [q_{11}, q_{12}, 0]$ and $q^2 \equiv [q_{21}, 0, q_{23}]$. The period t aggregate quantity level Q^t can be calculated directly using only information on q^t and the vector of quality adjustment factors, $\alpha \equiv [\alpha_1, \alpha_2, \alpha_3]$, or indirectly by deflating period t expenditure $v_t \equiv p^t \cdot q^t$ by the estimated period t price level, P^t. Thus we have the following two possible methods for constructing the Q^t:

$$(16)\quad Q^t \equiv \alpha \cdot q^t \;;\text{ or } Q^t \equiv p^t \cdot q^t/P^t \;; \qquad t = 1,2.$$

However, using the complete price vectors p^t with imputed prices filling in for the missing prices, equations (12) hold exactly and thus it is straightforward to show that $Q^t = \alpha \cdot q^t = p^t \cdot q^t/P^t$ for t = 1,2. Thus it does not matter whether we use the direct or indirect method for calculating the quantity levels; both methods give the same answer in this simple model.[30]

A problem with this simple model is that there is only one product that is present in both periods. In the following section, we generalize the present model to allow for multiple overlapping products.

4. A Nonstochastic Method for Quality Adjustment: A More Complex Model

In order to generalize the very simple model for dealing with new and disappearing products that was presented in the previous section, it is first necessary to develop another application of the fundamental equations (10) in section 2.

---

[28] Strictly speaking, it would be necessary to add a tiny amount to these prices to deter consumers from purchasing these products if they were made available.

[29] The use of carry forward and backward prices to estimate missing prices is widespread in statistical agencies. For additional materials on this method for estimating missing prices, see Triplett (2004), de Haan and Krsinich (2012) and Diewert, Fox and Schreyer (2017).

[30] In subsequent sections when we no longer assume that equations (12) hold exactly, then the direct and indirect methods for calculating the Q^t will in general differ.

Define the aggregator function Q(q) as follows:

(17) $Q_{KBF}(q^*) \equiv [q^*\cdot Aq^*]^{1/2}$

where q^* is defined as the N dimensional quantity vector $[q_1^*,...,q_N^*]$ and $A \equiv [a_{ij}]$ is an N by N symmetric matrix of parameters which satisfies certain regularity conditions.[31] Suppose further that that the observed price and quantity vectors for periods 1 and 2 are the positive price and quantity vectors, $p^{t*} \equiv [p_{t1}^*,...,p_{tN}^*]$ and $q^{t*} \equiv [q_{t1}^*,...,q_{tN}^*]$ for t = 1,2. We assume that q^{t*} solves $\max_q \{Q(q) : p^{t*}\cdot q = v^{t*} ; q \geq 0_N\}$ for t = 1,2 where $v^{t*} \equiv p^{t*}\cdot q^{t*}$ is observed expenditure on the N products for periods t = 1,2. The inverse demand functions (10) that correspond to this particular aggregator function are the following ones:

(18) $p^{t*} = P^{t*}\,\nabla_q Q_{KBF}(q^{t*}) = P^t\,[q^{t*}\cdot Aq^{t*}]^{-1/2}\,Aq^{t*}$; t = 1,2.

Using the framework described in section 2 above, the period t aggregate quantity level for the present model is $Q^{t*} \equiv [q^{t*}\cdot Aq^{t*}]^{1/2}$ and the corresponding period t price level is $P^{t*} \equiv p^{t*}\cdot q^{t*}/Q^{t*}$ for t = 1,2. The *Fisher* (1922) *ideal quantity index* is a function of the observable price and quantity data and is defined as follows:;

(19) $Q_F(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv [p^{1*}\cdot q^{2*}\ p^{2*}\cdot q^{2*}/p^{1*}\cdot q^{1*}\ p^{2*}\cdot q^{1*}]^{1/2}$.

Use equations (18) to eliminate p^{1*} and p^{2*} from the right hand side of (19). We find that

(20) $(p^{1*}\cdot q^{2*}\ p^{2*}\cdot q^{2*})/(p^{1*}\cdot q^{1*}\ p^{2*}\cdot q^{1*}) = q^{2*}\cdot Aq^{2*}/q^{1*}\cdot Aq^{1*}$.

Take positive square roots on both sides of (20). Using definitions (17) and (19), the resulting equation is:

(21) $Q_{KBF}(q^{2*})/Q_{KBF}(q^{1*}) = Q_F(p^{1*},p^{2*},q^{1*},q^{2*})$.

Thus $Q^{2*}/Q^{1*} = Q_{KBF}(q^{2*})/Q_{KBF}(q^{1*})$ is equal to the Fisher ideal quantity index $Q_F(p^{1*},p^{2*},q^{1*},q^{2*})$, which can be calculated using observable price and quantity data for the two periods. We know from section 2 that

(22) $P^{t*}Q^{t*} = p^{t*}\cdot q^{t*}$; t = 1,2.

Now make the normalization $P^{1*} = 1$. Using this normalization and equations (21) and (22), the aggregate price and quantity levels for the two periods can be defined in terms of observable data as follows:

[31] A is assumed to have one positive eigenvalue with a corresponding strictly positive eigenvector and N−1 negative or zero eigenvalues. This functional form was introduced into the economics literature by Konüs and Byushgens (1926) who showed its connection to the Fisher (1922) ideal index. This explains why $Q(q^*)$ is labeled as $Q_{KBF}(q^*)$. For further discussion of the regularity conditions on $Q_{KBF}(q^*)$, see Diewert (1976) and Diewert and Hill (2010).

(23) $P^{1*} \equiv 1$; $Q^{1*} \equiv p^{1*}\cdot q^{1*}$; $Q^{2*} \equiv Q^{1*}Q_F(p^{1*},p^{2*},q^{1*},q^{2*})$; $P^{2*} \equiv p^{1*}\cdot q^{1*}/Q^{2*}$.

The above results can be combined with the 3 product model that was described in the previous section: relabel the above aggregate data as a composite product 1 so that the new product 1 that corresponds to the first product in section 3 has prices and quantities defined as $p_{t1} \equiv P^{t*}$ and $q_{t1} \equiv Q^{t*}$ for t = 1,2. Products 2 and 3 are a disappearing product and a new product respectively as in section 3 above. The aggregate price levels for the two periods (which use all N+2 products) are P^1 and P^2 and the new α_n parameters are defined by the following counterparts to equations (14) above:

(24) $P^1 = 1$; $P^2 = P^{2*}/P^{1*} = P_F(p^{1*},p^{2*},q^{1*},q^{2*})$; $\alpha_1 = 1$; $\alpha_2 = p_{12}$; $\alpha_3 = p_{23}/(P^{2*}/P^{1*})$

where $P^{2*}/P^{1*} \equiv [v^{2*}/v^{1*}]/[Q^{2*}/Q^{1*}] \equiv P_F(p^{1*},p^{2*},q^{1*},q^{2*})$ is the Fisher (1922) ideal price index that compares the prices of the N products that are present in both periods, p^{1*}, p^{2*}, for the two periods under consideration. The imputed prices for the missing products, p_{13}^* and p_{22}^*, are obtained by using equations (15) for our present model:

(25) $p_{13}^* \equiv p_{23}/P_F(p^{1*},p^{2*},q^{1*},q^{2*})$; $p_{22}^* \equiv P_F(p^{1*},p^{2*},q^{1*},q^{2*})p_{12}$.

Comparing (24) and (25) with the corresponding equations (14) and (15) for the 3 product model, it can be seen that the price ratio for product 1 that was present in both periods, p_{21}/p_{11}, is replaced by the Fisher index $P_F(p^{1*},p^{2*},q^{1*},q^{2*})$ which is now defined over the set of products that are present in both periods. The type of inflation adjusted carry backward price p_{13}^* and the inflation adjusted carry forward price p_{22}^* defined by (25) are widely used by statistical agencies to estimate missing prices but agencies usually use Laspeyres or Paasche indexes in place of the Fisher price index.[32]

The aggregator function that is consistent with the new model with N continuing products, one disappearing product and one new product is defined as follows:

(26) $Q(q_1^*,...,q_N^*,q_2,q_3) \equiv \alpha_1 Q_{KBF}(q^*) + \alpha_2 q_2 + \alpha_3 q_3$

where $Q_{KBF}(q^*)$ is the KBF aggregator function defined by (17) and α_1 is set equal to 1.[33] Note that the model defined by (26) is restrictive from the economic perspective because the additive nature of definition (26) implies that the composite first commodity is perfectly substitutable with the new and disappearing commodities (which are also perfect substitutes for each other after quality adjustment). However, if the products under consideration *are* highly substitutable for each other, the implicit assumption of perfect substitutes for missing products will be acceptable. Moreover, the advantage of

[32] Note that the aggregate price index that is generated by this model is $P_F(p^{1*},p^{2*},q^{1*},q^{2*})$ which does not use the unmatched prices for the two periods.

[33] It is not necessary to use the KBF aggregator function in the above model; any aggregator function that has an exact index number associated with it will work. See Diewert (1976) for examples of exact index number formulae.

this form of quality adjustment is that it is relatively easy to explain to the public and it is fairly straightforward to implement.

The restriction that there is only one new product and one disappearing product is readily relaxed. The overall price index will continue to be $P_F(p^{1*},p^{2*},q^{1*},q^{2*})$ and counterparts to equations (25) can be used to generate imputed prices for the missing products.

We turn now to applications of the basic framework explained in section 2 where conditions (10) only hold approximately rather than exactly.

5. Time Product Dummy Regressions: The Case of No Missing Observations

In this section, it is assumed that price and quantity data for N products are available for T periods. Let $p^t \equiv [p_{t1},...,p_{tN}]$ and $q^t \equiv [q_{t1},...,q_{tN}]$ denote the price and quantity vectors for time periods t = 1,...,T. Initially, it is assumed that there are no missing prices or quantities so that all NT prices and quantities are positive. We assume that the quantity aggregator function Q(q) is the following linear function:

$$(27)\ Q(q) = Q(q_1,q_2,...,q_N) \equiv \Sigma_{n=1}^{N} \alpha_n q_n = \alpha \cdot q$$

where the α_n are positive parameters, which can be interpreted as quality adjustment factors. Under the assumption of maximizing behavior on the part of purchasers of the N commodities, assumption (27) applied to equations (10) imply that the following NT equations should hold exactly:

$$(28)\ p_{tn} = \pi_t \alpha_n \ ; \qquad n = 1,...,N;\ t = 1,...,T$$

where we have redefined the period t price levels P^t in equations (10) as the parameters π_t for t = 1,...,T.

Note that equations (28) form the basis for the *time dummy hedonic regression model* which is due to Court (1939).[34] It can be seen that these equations are a special case of the general model of consumer behavior that was explained in section 2 above.

At this point, it is necessary to point out that our consumer theory derivation of equations (28) is not accepted by all economists. Rosen (1974), Triplett (1987) (2004) and Pakes (2001)[35] have argued for a more general approach to the derivation of hedonic regression models that is based on supply conditions as well as on demand conditions. The present

[34] This was Court's (1939; 109-111) hedonic suggestion number two. He transformed the underlying equations (28) by taking logarithms of both sides of these equations (which will be done below). He chose to transform the prices by the log transformation because the resulting regression model fit his data on automobiles better. Diewert (2003b) also recommended the log transformation on the grounds that multiplicative errors were more plausible than additive errors.

[35] "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).

approach is obviously based on consumer demands and preferences only. This consumer oriented approach was endorsed by Griliches (1971; 14-15), Muellbauer (1974; 988) and Diewert (2003a) (2003b).[36] Of course, the separability assumptions which justify the present consumer approach are quite restrictive but nevertheless, it is useful to imbed hedonic regression models in a traditional consumer demand setting.

Empirically, equations (28) are unlikely to hold exactly. Thus we assume that the exact model defined by (28) holds only to some degree of approximation and so error terms, e_{tn}, are added to the right hand sides of equations (28). The unknown parameters, $\pi \equiv [\pi_1,...,\pi_T]$ and $\alpha \equiv [\alpha_1,...,\alpha_N]$, will be estimated as solutions to the following (nonlinear) least squares minimization problem:

$$(29)\ \min_{\alpha,\pi} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} [p_{tn} - \pi_t\alpha_n]^2 .$$

Our approach to the specification of the error terms will not be very precise. Throughout the paper, we will obtain estimators for the aggregate price levels π_t and the quality adjustment parameters α_n as solutions to least squares minimization problems like those defined by (29) or as solutions to weighted least squares minimization problems that will be considered in subsequent sections. Our focus will not be on the distributional aspects of our estimators; rather, our focus will be on the *axiomatic* or *test properties* of the price levels that are solutions to the various least squares minimization problems.[37] Basically, the approach taken here is a descriptive statistics approach: we consider simple models that aggregate price and quantity information for a given period over a set of specified commodities into scalar measures of aggregate price and quantity that summarize the detailed price and quantity information in a "sensible" way.[38]

The first order necessary (and sufficient) conditions for $\pi \equiv [\pi_1,...,\pi_T]$ and $\alpha \equiv [\alpha_1,...,\alpha_N]$ to solve the minimization problem defined by (29) are equivalent to the following N + T equations:

---

[36] Diewert (2003b; 97) justified the consumer demand approach as follows: "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. 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 an 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." Footnote 25 on page 82 of Diewert (2003b) explains how the present hedonic model can be derived from Diewert's (2003a) consumer based model by strengthening the assumptions in the 2003a paper.

[37] For rigorous econometric approaches to the stochastic approach to index number theory, see Rao and Hajargasht (2016) and Gorajek (2018). These papers consider many transformations of the fundamental hedonic equations (28) and many methods for constructing averages of prices.

[38] Our approach is broadly similar to Theil's (1967; 136-137) approach to index number theory.

(30) $\alpha_n = \Sigma_{t=1}^{T} \pi_t p_{tn}/\Sigma_{t=1}^{T} \pi_t^2$ $\qquad$ n = 1,...,N
$\quad = \Sigma_{t=1}^{T} \pi_t^2 (p_{tn}/\pi_t)/\Sigma_{t=1}^{T} \pi_t^2$;
(31) $\pi_t = \Sigma_{n=1}^{N} \alpha_n p_{tn}/\Sigma_{n=1}^{N} \alpha_n^2$ $\qquad$ t = 1,...,T
$\quad = \Sigma_{n=1}^{N} \alpha_n^2(p_{tn}/\alpha_n)/\Sigma_{n=1}^{N} \alpha_n^2$.

Solutions to the two sets of equations can readily be obtained by iterating between the two sets of equations. Thus set $\alpha^{(1)} = 1_N$ (a vector of ones of dimension N) in equations (31) and calculate the resulting $\pi^{(1)} = [\pi_1^{(1)},...,\pi_T^{(1)}]$. Then substitute $\pi^{(1)}$ into the right hand sides of equations (30) to calculate $\alpha^{(2)} \equiv [\alpha_1^{(2)},...,\alpha_N^{(2)}]$. And so on until convergence is achieved.

If $\pi^* \equiv [\pi_1^*,...,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,...,\alpha_N^*]$ is a solution to (30) and (31), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$. Then $1,\pi_2^*,...,\pi_T^*$ is the sequence of fixed base index numbers that is generated by the least squares minimization problem defined by (29).

If quantity data are available, then using the general methodology that was outlined in section 2, aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^*\cdot q^t = \Sigma_{n=1}^{N} \alpha_n^* q_{tn}$ for t = 1,...,T. Estimated aggregate price levels can be obtained directly from the solution to (29); i.e., set $P^{t*} = \pi_t^*$ for t = 1,...,T. Alternative price levels can be indirectly obtained as $P^{t**} \equiv p^t\cdot q^t/Q^{t*} = p^t\cdot q^t/\alpha^*\cdot q^t$ for t = 1,...,T. If the optimized objective function in (29) is 0 (so that all errors $e_{tn}^* \equiv p_{tn} - \pi_t^*\alpha_n^*$ equal 0 for t = 1,...,T and n = 1,...,N), then P^{t*} will equal P^{t**} for all t. However, usually nonzero errors will occur and so a choice between the two sets of estimators must be made.[39]

From (30), it can be seen that α_n^*, the quality adjustment parameter for product n, is a weighted average of the T inflation adjusted prices for product n, the p_{tn}/π_t^*, where the weight for p_{tn}/π_t^* is $\pi_t^{*2} /\Sigma_{\tau=1}^{T} \pi_\tau^{*2}$. This means that the weight for p_{tn}/π_t^* will be very high for periods t where general inflation is high, which seems rather arbitrary. From (31), it can be seen that π_t^*, the period t price level (and fixed base price index), is weighted average of the N quality adjusted prices for period t, the p_{tn}/α_n^*, where the weight for p_{tn}/α_n^* is $\alpha_n^{*2} /\Sigma_{i=1}^{N} \alpha_i^{*2}$. It is a positive feature of the method that π_t^* is a weighted average of the quality adjusted prices for period t but the quadratic nature of the weights is not an attractive feature.

In addition to having unattractive weighting properties, the estimates generated by solving the least squares minimization problem (29) suffer from a fatal flaw: *the estimates are not invariant to changes in the units of measurement*. In order to remedy this defect, we turn to an alternative error specification.

[39] Usually, the direct estimates for the price levels will be used in hedonic regression studies; i.e., the $P^{t*} = \pi_t^*$ estimates will be used. For statistical agencies, an advantage of the direct estimates is that they can be calculated without the use of quantity information. However, later in this paper, we will note some advantages of the indirect method if quantity information is available.

Instead of adding approximation errors to the exact equations (28), we could append multiplicative approximation errors. Thus the exact equations become $p_{tn} = \pi_t\alpha_n e_{tn}$ for n = 1,...,N and t = 1,...,T. Upon taking logarithms of both sides of these equations, we obtain the following system of estimating equations:

$$(32)\ \ln p_{tn} = \ln\pi_t + \ln\alpha_n + \ln e_{tn}\ ; \qquad n = 1,...,N;\ t = 1,...,T$$
$$= \rho_t + \beta_n + \varepsilon_{tn}$$

where $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and $\beta_n \equiv \ln\alpha_n$ for n = 1,...,N. The model defined by (32) is the basic *Time Product Dummy regression model* with no missing observations.[40] Now choose the ρ_t and β_n to minimize the sum of squared residuals, $\Sigma_{n=1}^{N} \Sigma_{t=1}^{T} \varepsilon_{tn}^{2}$. Thus let $\rho \equiv [\rho_1,...,\rho_T]$ and $\beta \equiv [\beta_1,...,\beta_N]$ be a solution to the following least squares minimization problem:

$$(33)\ \min_{\rho,\,\beta} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} [\ln p_{tn} - \rho_t - \beta_n]^2\ .$$

The first order necessary conditions for $\rho_1,...,\rho_T$ and $\beta_1,...,\beta_N$ to solve (33) are the following T + N equations:

$$(34)\ N\rho_t + \Sigma_{n=1}^{N} \beta_n = \Sigma_{n=1}^{N} \ln p_{tn}\ ; \qquad t = 1,...,T;$$
$$(35)\ \Sigma_{t=1}^{T} \rho_t + T\beta_n = \Sigma_{t=1}^{T} \ln p_{tn}\ ; \qquad n = 1,...,N.$$

Replace the ρ_t and β_n in equations (34) and (35) by $\ln\pi_t$ and $\ln\alpha_n$ respectively for t = 1,...,T and n = 1,...,N. After some rearrangement, the resulting equations become:

$$(36)\ \pi_t = \Pi_{n=1}^{N} (p_{tn}/\alpha_n)^{1/N}\ ; \qquad t = 1,...,T;$$
$$(37)\ \alpha_n = \Pi_{t=1}^{T} (p_{tn}/\pi_t)^{1/T}\ ; \qquad n = 1,...,N.$$

Thus the period t aggregate price level, π_t, is equal to the geometric average of the N quality adjusted prices for period t, p_{t1}/α_1, ..., p_{tN}/α_N, while the quality adjustment factor for product n, α_n, is equal to the geometric average of the T inflation adjusted prices for product n, p_{1n}/π_1, ..., p_{Tn}/π_T. These estimators look very reasonable (if quantity weights are not available).

Solutions to (36) and (37) can readily be obtained by iterating between the two sets of equations. Thus set $\alpha^{(1)} = 1_N$ (a vector of ones of dimension N) in equations (36) and calculate the resulting $\pi^{(1)} = [\pi_1^{(1)},...,\pi_T^{(1)}]$. Then substitute $\pi^{(1)}$ into the right hand sides of equations (37) to calculate $\alpha^{(2)} \equiv [\alpha_1^{(2)},...,\alpha_N^{(2)}]$. And so on until convergence is achieved. Alternatively, equations (34) and (35) are linear in the unknown parameters and can be solved (after normalizing one parameter) by a simple matrix inversion. A final method of

[40] In the statistics literature, this type of model is known as a fixed effects model. A generalized version of this model (with missing observations) was proposed by Summers (1973) in the international comparison context where it is known as the Country Product Dummy regression model. A weighted version of this model (with missing observations) was proposed by Aizcorbe, Corrado and Doms (2000).

obtaining a solution to (34) and (35) is to apply a simple linear regression model to equations (32).[41]

If $\pi^* \equiv [\pi_1^*,...,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,...,\alpha_N^*]$ is a solution to (36) and (37), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$ (which corresponds to $\rho_1 = 0$). Then $1,\pi_2^*,...,\pi_T^*$ is the sequence of fixed base index numbers that is generated by the least squares minimization problem defined by (33).

Once we have the unique solution $1,\pi_2^*,...,\pi_T^*$ for the T price levels that are generated by the (33), the *price index* between period t relative to period s can be defined as π_t^*/π_s^*. Using equations (36) for π_t^* and π_s^*, we have the following expression for these price indexes:

$$(38)\quad \pi_t^*/\pi_s^* = \Pi_{n=1}^N (p_{tn}/\alpha_n^*)^{1/N}/\ \Pi_{n=1}^N (p_{sn}/\alpha_n^*)^{1/N} = \Pi_{n=1}^N (p_{tn}/p_{sn})^{1/N}.$$

Thus if there are no missing observations, the Time Product Dummy price indexes between any two periods in the window of T period under consideration is equal to the *Jevons index* between the two periods (the simple geometric mean of the price ratios, p_{tn}/p_{sn}).[42] This is a somewhat disappointing result since an equally weighted average of the price ratios is not necessarily a representative average of the prices; i.e., unimportant products to purchasers (in the sense that they spend very little on these products) are given the same weight in the Jevons measure of inflation between the two periods as is given to high expenditure products.[43]

If there are no missing observations, then it can be seen using equations (37) that the ratio of the quality adjustment factor for product n relative to product m is equal to the following sensible expression:

$$(39)\quad \alpha_n^*/\alpha_m^* = \Pi_{t=1}^T (p_{tn}/\pi_t^*)^{1/T} / \Pi_{t=1}^T (p_{tm}/\pi_t^*)^{1/T} = \Pi_{t=1}^T (p_{tn}/p_{tm})^{1/T}.$$

If quantity data are available, then aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^*\cdot q^t = \Sigma_{n=1}^N \alpha_n^* q_{tn}$ for t = 1,...,T. Estimated aggregate price levels can be obtained directly from the solution to (33); i.e., set $P^{t*} = \pi_t^*$ for t = 1,...,T. Alternative price levels can be obtained indirectly as $P^{t**} \equiv p^t\cdot q^t/Q^{t*} = p^t\cdot q^t/\alpha^*\cdot q^t$ for t = 1,...,T.[44] If the optimized objective function in (33) is 0 (so that all errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ equal 0 for t = 1,...,T and n = 1,...,N), then P^{t*} will equal P^{t**} for all t. If the estimated residuals

[41] Again we require one normalization on the parameters such as $\rho_1 = 0$.
[42] This result is a special case of a more general result obtained by Triplett and McDonald (1977; 150).
[43] However, if quantity data are not available, the Jevons index has the strongest axiomatic properties compared to its competitors; see the ILO, Eurostat, IMF, OECD, UNECE and the World Bank (2004).
[44] The fact that a time dummy hedonic regression model generates two alternative decompositions of the value aggregate into price and quantity aggregates was noted in de Haan and Krsinich (2018).

are not all equal to 0, then the two estimates for the period t price level P^t will differ. The two estimates for P^t will generate different estimates for the companion aggregate quantity levels.

Note that the underlying exact model ($p_{tn} = \pi_t\alpha_n$ for t and n) is the same for both least squares minimization problems, (29) and (33). However, different error specifications and different transformations of both sides of the equations $p_{tn} = \pi_t\alpha_n$ can lead to very different estimators for the π_t and α_n. Our strategy in this section and in the following sections will be to choose specifications of the least squares minimization problem that lead to estimators for the price levels π_t that have good axiomatic properties.[45] From this perspective, it is clear that (33) leads to "better" estimates than (29).

In the following section, we allow for missing observations.

6. Time Product Dummy Regressions: The Case of Missing Observations

In this section, the least squares minimization problem defined by (33) is generalized to allow for missing observations. In order to make this generalization, it is first necessary to make some definitions. As in the previous section, there are N products and T time periods but not all products are purchased (or sold) in all time periods. For each period t, define the set of products n that are present in period t as $S(t) \equiv \{n: p_{tn} > 0\}$ for t = 1,2,...,T. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t: p_{tn} > 0\}$. Again, assume that these sets are not empty; i.e., each product is sold in at least one time period. Define the integers N(t) and T(n) as follows:

$$(40)\quad N(t) \equiv \Sigma_{n\in S(t)}\, 1; \qquad t = 1,...,T;$$
$$(41)\quad T(n) \equiv \Sigma_{t\in S^*(n)}\, 1; \qquad n = 1,...,N.$$

If all N products are present in period t, then N(t) = N; if product n is present in all T periods, then T(n) = T.

Using the notation that was defined in the previous section, the counterpart to (33) when there are missing products is the following least squares minimization problem:

$$(42)\quad \min_{\rho,\beta} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho,\beta} \Sigma_{n=1}^{N} \Sigma_{t\in S^*(n)} [\ln p_{tn} - \rho_t - \beta_n]^2.$$

Note that there are two equivalent ways of writing the least squares minimization problem.[46] The first order necessary conditions for $\rho_1,...,\rho_T$ and $\beta_1,...,\beta_N$ to solve (42) are the following counterparts to (34) and (35):

---

[45] From the perspective of the economic approach to index number theory that was outlined in section 2, problems (29) and (33) have the same economic justification.

[46] The first expression is used when (42) is differentiated with respect to ρ_t and the second expression is used when differentiating (42) with respect to β_n.

(43) $\Sigma_{n\in S(t)} [\rho_t + \beta_n] = \Sigma_{n\in S(t)} \ln p_{tn}$; t = 1,...,T;
(44) $\Sigma_{t\in S^*(n)} [\rho_t + \beta_n] = \Sigma_{t\in S^*(n)} \ln p_{tn}$; n = 1,...,N.

As in the previous section, let $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and let $\beta_n \equiv \ln\alpha_n$ for n = 1,...,N. Substitute these definitions into equations (43) and (44). After some rearrangement and using definitions (40) and (41), equations (43) and (44) become the following ones:

(45) $\pi_t = \Pi_{n\in S(t)} [p_{tn}/\alpha_n]^{1/N(t)}$; t = 1,...,T;
(46) $\alpha_n = \Pi_{t\in S^*(n)} [p_{tn}/\pi_t]^{1/T(n)}$; n = 1,...,N.

The same iterative procedure that was explained in the previous section will work to generate a solution to equations (45) and (46).[47] As was the case in the previous section, solutions to (45) and (46) are not unique; if π^*, α^* is a solution to (45) and (46), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$ (which corresponds to $\rho_1 = 0$). Then $1,\pi_2^*,...,\pi_T^*$ is the sequence of (normalized) price levels that is generated by the least squares minimization problem defined by (42).[48] In this case, $\pi_t^* = \Pi_{n\in S(t)}[p_{tn}/\alpha_n^*]^{1/N(t)}$ is the equally weighted geometric mean of all of the quality adjusted prices for the products that are available in period t for t = 2,3,...,T and the quality adjustment factors are normalized so that $\pi_1^* = \Pi_{n\in S(1)} [p_{1n}/\alpha_n^*]^{1/N(1)} = 1$. From (46), we can deduce that α_n^* will be larger for products that are relatively expensive and will be smaller for cheaper products.

Once we have the unique solution $1,\pi_2^*,...,\pi_T^*$ for the T price levels that are generated by the (42), the *price index* between period t relative to period r can be defined as π_t^*/π_r^*. Using equations (45) and (46), we have the following expressions for π_t^*/π_r^* and α_n^*/α_m^*:

(47) $\pi_t^*/\pi_r^* = \Pi_{n\in S(t)}[p_{tn}/\alpha_n^*]^{1/N(t)} / \Pi_{n\in S(r)}[p_{rn}/\alpha_n^*]^{1/N(r)}$; $1 \leq t, r \leq T$;
(48) $\alpha_n^*/\alpha_m^* = \Pi_{t\in S^*(n)}[p_{tn}/\pi_t^*]^{1/T(n)} / \Pi_{t\in S^*(m)}[p_{tm}/\pi_t^*]^{1/T(m)}$; $1 \leq n, m \leq N$.

Note that in general, the quality adjustment factors α_n^* do not cancel out for the indexes π_t^*/π_r^* defined by (47) as they did in the previous section. However, these price indexes do have some good axiomatic properties.[49] If the set of available products is the same in

[47] Of course, it is not necessary to use the iterative procedure to find a solution to equations (43) and (44). After setting $\rho_1 = 0$ and dropping the first equation in (43), matrix algebra can be used to find a solution to the remaining equations. Alternatively, after setting $\rho_1 = 0$, use the equations $\ln p_{tn} = \rho_t + \beta_n + \varepsilon_{tn}$ for t = 1,...,T and $n\in S(t)$ to set up a linear regression model with time and product dummy variables and use a standard ordinary least squares econometric software package to obtain the solution $\rho_1^* = 0$, $\rho_2^*,...,\rho_T^*$, $\beta_1^*,...,\beta_N^*$ to (42).
[48] We need enough observations on products that are present so that a full rank condition is satisfied for the modified version of equations (43) and (44). If there is a rapid proliferation of new and disappearing products, then it may not be possible to invert the coefficient matrix that is associated with the modified equations (43) and (44). In subsequent models with missing observations, we will assume that a similar full rank condition is satisfied.
[49] The index π_t^*/π_r^* satisfies the identity test (if prices are the same in periods r and t, then the index is equal to 1) and it is invariant to changes in the units of measurement. It is also homogeneous of degree one in the prices of period t and homogeneous of degree minus one in the prices of period r.

periods r and t, then the quality adjustment factors do cancel and the price index for period t relative to period r is $\pi_t^*/\pi_r^* = \Pi_{n\in S(t)}[p_{tn}/p_{rn}]^{1/N(t)}$, which is the Jevons index between periods r and t. Again, while this index is a excellent one if quantity information is not available, it is not satisfactory when quantity information is available due to its equal weighting of economically important and unimportant price ratios.[50]

There is another unsatisfactory property of the estimated price levels that are generated by solving the time product dummy hedonic model that is defined by (42): a product that is available only in one period out of the T periods has no influence on the aggregate price levels π_t^*.[51] To see this, consider equations (43) and (44) and suppose that product n^* was only available in period t^*.[52] Equation n^* in N equations in (44) becomes the equation: $[\rho_{t^*} + \beta_{n^*}] = \ln p_{t^*n^*}$. Thus once ρ_{t^*} has been determined, β_{n^*} can be defined as $\beta_{n^*} \equiv \ln p_{t^*n^*} - \rho_{t^*}$. Subtract the equation $[\rho_{t^*} + \beta_{n^*}] = \ln p_{t^*n^*}$ from equation t^* and the resulting equations in (43) can be written as equations (49) below. Dropping equation n^* in equations (44) leads to equations (50) below:

(49) $\Sigma_{n\in S(t),\, n\neq n^*} [\rho_t + \beta_n] = \Sigma_{n\in S(t),\, n\neq n^*} \ln p_{tn}$; $\qquad t = 1,...,T;$

(50) $\Sigma_{t\in S^*(n)} [\rho_t + \beta_n] = \Sigma_{t\in S^*(n)} \ln p_{tn}$; $\qquad n = 1,...,n^*-1,n^*+1,...,N.$

Equations (49) and (50) are T+N−1 equations which do not involve $p_{t^*n^*}$. After making the normalization $\rho_1^* = 0$, these equations can be solved for $\rho_2^*,..., \rho_T^*, \beta_1^*,..., \beta_{n^*-1}^*, \beta_{n^*+1}^*,..., \beta_N^*$. Now define $\beta_{n^*}^* \equiv \ln p_{t^*n^*} - \rho_{t^*}$ and we have the (normalized) solution for (42). Since the ρ_t^* do not involve $p_{t^*n^*}$, the resulting $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,...,T also do not depend on the isolated price $p_{t^*n^*}$. This proof can be repeated for any number of isolated prices. This property of the time product dummy model is unfortunate because it means that when a new product enters the marketplace in period T, it has no influence on the price levels $1,\pi_2^*,...,\pi_T^*$ that are generated by solving the least squares minimization problem defined by (42).

If quantity data are available, then aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \Sigma_{n\in S(t)} \alpha_n^* q_{tn}$ for t = 1,...,T.[53] Estimated aggregate price levels can be obtained directly from the solution to (42); i.e., set $P^{t*} = \pi_t^*$ for t = 1,...,T. Alternative price levels can be obtained indirectly as $P^{t**} \equiv \Sigma_{n\in S(t)} p_n q_{tn}/Q^{t*} = \Sigma_{n\in S(t)} p_n q_{tn}/\Sigma_{n\in S(t)} \alpha_n^* q_{tn}$ for t = 1,...,T.[54] If the optimized objective function in (42) is 0, so that all errors $\varepsilon_{tn}^* \equiv \ln$

[50] However, if the estimated squared residuals are small in magnitude for periods τ and t, then the index π_t^*/π_r^* defined by (47) will be satisfactory since in this case $p^\tau \approx \pi_\tau^*\alpha^*$ and $p^t \approx \pi_t^*\alpha^*$ so that prices are approximately proportional for these two periods and π_t^*/π_r^* defined by (47) will be approximately correct. Any missing prices for any period t and product n are defined as $p_{tn}^* \equiv \pi_t^*\alpha_n^*$.

[51] This property of the Time Product Dummy model was first noticed by Diewert (2004) (in the context of the Country Product Dummy model).

[52] We assume that products other than product n^* are available in period t^*.

[53] Note that each $\alpha_n^* > 0$ since $\alpha_n^* \equiv \exp[\beta_n^*]$ for n = 1,...,N.

[54] Note that $P^{t**} \equiv \Sigma_{n\in S(t)} p_{tn}q_{tn}/\Sigma_{n\in S(t)} \alpha_n^* q_{tn}$ is a period t *quality adjusted unit value price level*. The corresponding quantity level is $Q^{t**} \equiv \Sigma_{n\in S(t)} p_{tn}q_{tn}/P^{t**} = \Sigma_{n\in S(t)} \alpha_n^* q_{tn}$ which is a *linear aggregator function*. Quality adjusted unit value price levels and indexes were introduced by Dalén (2001) and further studied by de Haan (2004b) (2010), von Auer (2014) (for the case of no missing observations) and de Haan and

$p_{tn} - \rho_t^* - \beta_n^*$ equal 0 for t = 1,...,T and $n \in S(t)$, then P^{t*} will equal P^{t**} for all t. If the estimated residuals are not all equal to 0, then the two estimates for the period t price level P^t will differ. The two estimates for P^t will generate different estimates for the companion aggregate quantity levels.

7. Weighted Time Product Dummy Regressions: The Bilateral Case

A major problem with the indexes discussed in the previous 2 sections is the fact that they do not weight the individual product prices by their economic importance. The first serious index number economist to stress the importance of weighting was Walsh (1901).[55] Keynes was quick to follow up on the importance of weighting[56] and Fisher emphatically endorsed weighting.[57] Griliches also endorsed weighting in the hedonic regression context.[58]

Krsinich (2018). By looking at (42), it can be seen that if prices are identical in periods t and r so that $p^t = p^r$, then $P^{t*} = P^{r*}$; i.e., an identity test for the direct hedonic price levels will be satisfied. However, the corresponding Q^{t*} will not satisfy the identity test for quantity levels; i.e., if quantities q_{tn} and q_{rn} are equal in periods t and r, it is not the case that $Q^{t*} \equiv \Sigma_{n=1}^N p_{tn}q_{tn}/\pi_t^*$ will equal $Q^{r*} \equiv \Sigma_{n=1}^N p_{trn}q_{rn}/\pi_r^*$ for $r \neq t$. On the other hand, it can be seen that $Q^{t**} = \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn} = \Sigma_{n\in S(t)}\, \alpha_n^* q_{rn} = Q^{r*}$ if $q_{tn} = q_{rn}$ for all n. Thus the choice between using P^{t*} or P^{t**} could be made on the basis of choosing which identity test is more important to satisfy. The analysis here follows that of de Haan and Krsinich (2018; 763-764)

[55] See Walsh (1901). This book laid the groundwork for the test or axiomatic approach to index number theory which was further developed by Fisher (1922). In his second book on index number theory, Walsh made the case for weighting by economic importance as follows: “It might seem at first sight as if simply every price quotation were a single item, and since every commodity (any kind of commodity) has one price-quotation attached to it, it would seem as if price-variations of every kind of commodity were the single item in question. This is the way the question struck the first inquirers into price-variations, wherefore they used simple averaging with even weighting. But a price-quotation is the quotation of the price of a generic name for many articles; and one such generic name covers a few articles, and another covers many. … A single price-quotation, therefore, may be the quotation of the price of a hundred, a thousand, or a million dollar’s worths, of the articles that make up the commodity named. Its weight in the averaging, therefore, ought to be according to these money-unit’s worth.” Correa Moylan Walsh (1921a; 82-83).

[56] “It is also clear that the so-called unweighted index numbers, usually employed by practical statisticians, are the worst of all and are liable to large errors which could have been easily avoided.” J.M. Keynes (1909; 79). This paper won the Cambridge University Adam Smith Prize for that year. Keynes (1930; 76-77) again stressed the importance of weighting in a later paper which drew heavily on his 1909 paper.

[57] “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).

[58] “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.” Zvi Griliches (1971; 8).

In this section, we will discuss some alternative methods for weighting by economic importance in the context of a bilateral time product dummy regression model.[59] In this section, we will assume that there are no missing observations.

Recall the least squares minimization problem defined by (33) in section 5 above. The squared residuals, $[\ln p_{tn} - \rho_t - \beta_n]^2$, appear in this problem without any weighting. Thus products which have a high volume of sales in any period are given the same weight in the least squares minimization problem as products which have very few sales. In order to take economic importance into account, for the case of 2 time periods, replace (33) by the following *weighted least squares minimization problem*:

$$(51)\ \min_{\rho,\beta} \Sigma_{n=1}^N q_{1n}[\ln p_{1n} - \beta_n]^2 + \Sigma_{n=1}^N q_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2$$

where we have set $\rho_1 = 0$. The squared error for product n in period t is repeated q_{tn} times to reflect the sales of the product in period t. Thus the new problem (51) takes into account the popularity of each product.[60]

The first order necessary conditions for the minimization problem defined by (51) are the following N + 1 equations:

$$(52)\ (q_{1n}+q_{2n})\beta_n = q_{1n}\ln p_{1n} + q_{2n}(\ln p_{2n} - \rho_2)\ ; \qquad n = 1,...,N;$$
$$(53)\ (\Sigma_{n=1}^N q_{2n})\rho_2 = (\Sigma_{n=1}^N q_{2n})(\ln p_{2n} - \beta_n).$$

The solution to (52) and (53) is the following one:[61]

$$(54)\ \rho_2^* \equiv \Sigma_{n=1}^N q_{1n}q_{2n}(q_{1n} + q_{2n})^{-1} \ln(p_{2n}/p_{1n})/\Sigma_{i=1}^N q_{1i}q_{2i}(q_{1i} + q_{2i})^{-1};$$
$$(55)\ \beta_n^* \equiv q_{1n}(q_{1n} + q_{2n})^{-1} \ln(p_{1n}) + q_{2n}(q_{1n} + q_{2n})^{-1} \ln(p_{2n}/\pi_2^*)\ ; \qquad n = 1,...,N$$

where $\pi_2^* \equiv \exp[\rho_2^*]$. Note that the weight for the term $\ln(p_{2n}/p_{1n})$ in (54) can be written as follows:

$$(56)\ q_n^* \equiv \Sigma_{n=1}^N q_{1n}q_{2n}(q_{1n} + q_{2n})^{-1} /\Sigma_{i=1}^N q_{1i}q_{2i}(q_{1i} + q_{2i})^{-1}\ ; \qquad n = 1,...,N$$
$$= h(q_{1n},q_{2n})/\Sigma_{i=1}^N h(q_{1i},q_{2i})$$

where $h(a,b) \equiv 2ab/(a+b) = [½\, a^{-1} + ½\, b^{-1}]^{-1}$ is the *harmonic mean* of a and b.[62]

[59] The approach taken in this section is based on Rao (1995) (2004) (2005) and Diewert (2003b), (2005a) (2005b). Diewert (2005a) considered all four forms of weighting that will be discussed in this section while Rao (1995) (2005) discussed mainly the third form of weighting.

[60] One can think of repeating the term $[\ln p_{1n} - \beta_n]^2$ for each unit of product n sold in period 1. The result is the term $q_{1n}[\ln p_{1n} - \beta_n]^2$. A similar justification based on repeating the price according to its sales can also be made. This repetition methodology makes the stochastic specification of the error terms somewhat complicated. However, as indicated in the introduction, we leave these difficult distributional problems to other more capable econometricians.

[61] See Diewert (2005a).

[62] h(a,b) is well defined by ab/(a+b) if a and b are nonnegative and at least one of these numbers is positive. In order to write h(a,b) as $[½\, a^{-1} + ½\, b^{-1}]^{-1}$, we require a > 0 and b > 0.

Note that the q_n^* sum to 1 and thus ρ_2^* is a weighted average of the logarithmic price ratios $\ln(p_{2n}/p_{1n})$. Using $\pi_2^* = \exp[\rho_2^*]$ and $\pi_1^* = \exp[\rho_1^*] = \exp[0] = 1$, the bilateral price index that is generated by the solution to (51) is

(57) $\pi_2^*/\pi_1^* = \exp[\rho_2^*] = \exp[\Sigma_{n=1}^N q_n^* \ln(p_{2n}/p_{1n})]$.

Thus π_2^*/π_1^* is a weighted geometric mean of the price ratios p_{2n}/p_{1n} with weights q_n^* defined by (56). Although this seems to be a reasonable bilateral index number formula, it must be rejected for practical use on the grounds that *the index is not invariant to changes in the units of measurement*.

Since values are invariant to changes in the units of measurement, the lack of invariance problem could be solved if we replace the quantity weights in (51) with expenditure or sales weights.[63] This leads to the following weighted least squares minimization problem where the weights v_{tn} are defined as $p_{tn}q_{tn}$ for t = 1,2 and n = 1,...,N:

(58) $\min_{\rho,\beta} \Sigma_{n=1}^N v_{1n}[\ln p_{1n} - \beta_n]^2 + \Sigma_{n=1}^N v_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2$.

It can be seen that problem (58) has exactly the same mathematical form as problem (51) except that v_{tn} has replaced q_{tn} and so the solutions (54) and (55) will be valid in the present context if v_{tn} replaces q_{tn} in these formulae. Thus the solution to (58) is:

(59) $\rho_2^* \equiv \Sigma_{n=1}^N v_{1n}v_{2n}(v_{1n} + v_{2n})^{-1} \ln(p_{2n}/p_{1n})/\Sigma_{i=1}^N v_{1i}v_{2i}(v_{1i} + v_{2i})^{-1}$;
(60) $\beta_n^* \equiv v_{1n}(v_{1n} + v_{2n})^{-1} \ln(p_{1n}) + v_{2n}(v_{1n} + v_{2n})^{-1} \ln(p_{2n}/\pi_2^*)$; n = 1,...,N

where $\pi_2^* \equiv \exp[\rho_2^*]$.

The resulting price index, $\pi_2^*/\pi_1^* = \pi_2^* = \exp[\rho_2^*]$ is indeed invariant to changes in the units of measurement. However, if we regard π_2^* as a function of the price and quantity vectors for the two periods, say $P(p^1,p^2,q^1,q^2)$, then another problem emerges for the price index defined by the solution to (58): $P(p^1,p^2,q^1,q^2)$ is not homogeneous of degree 0 in the components of q^1 or in the components of q^2. These properties are important because it is desirable that the companion implicit quantity index defined as $Q(p^1,p^2,q^1,q^2) \equiv [p^2 \cdot q^2/p^1 \cdot q^1]/P(p^1,p^2,q^1,q^2)$ be homogeneous of degree 1 in the components of q^2 and homogeneous of degree minus 1 in the components of q^1.[64] We also want $P(p^1,p^2,q^1,q^2)$ to

---

[63] "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).

[64] Thus we want Q to have the following properties: $Q(p^1,p^2,q^1,\lambda q^2) = \lambda Q(p^1,p^2,q^1,q^2)$ and $Q(p^1,p^2,\lambda q^1,q^2) = \lambda^{-1}Q(p^1,p^2,q^1,q^2)$ for all $\lambda > 0$. For a list of desirable properties or tests for bilateral price indexes of the form $P(p^1,p^2,q^1,q^2)$, see Diewert (1992) or the ILO, Eurostat, IMF, OECD, UNECE and the World Bank (2004).

be homogeneous of degree 1 in the components of p^2 and homogeneous of degree minus 1 in the components of p^1 and these properties are also not satisfied. Thus we conclude that the solution to the weighted least squares problem defined by (58) does not generate a satisfactory price index formula.

The above deficiencies can be remedied if the *expenditure amounts* v_{tn} in (58) are replaced by *expenditure shares*, s_{tn}, where $v_t \equiv \Sigma_{n=1}^N v_{tn}$ for t = 1,2 and $s_{tn} \equiv v_{tn}/v_t$ for t = 1,2 and n = 1,...,N. This replacement leads to the following weighted least squares minimization problem:[65]

$$(61)\ \min_{\rho,\beta} \Sigma_{n=1}^N s_{1n}[\ln p_{1n} - \beta_n]^2 + \Sigma_{n=1}^N s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2.$$

Again, it can be seen that problem (61) has exactly the same mathematical form as problem (51) except that s_{tn} has replaced q_{tn} and so the solutions (54) and (55) will be valid in the present context if s_{tn} replaces q_{tn} in these formulae. Thus the solution to (61) is:

$$(62)\ \rho_2^* \equiv \Sigma_{n=1}^N s_{1n}s_{2n}(s_{1n} + s_{2n})^{-1} \ln(p_{2n}/p_{1n})/\Sigma_{i=1}^N s_{1i}s_{2i}(s_{1i} + s_{2i})^{-1};$$

$$(63)\ \beta_n^* \equiv s_{1n}(s_{1n} + s_{2n})^{-1} \ln(p_{1n}) + s_{2n}(s_{1n} + s_{2n})^{-1} \ln(p_{2n}/\pi_2^*)\ ; \qquad n = 1,...,N$$

where $\pi_2^* \equiv \exp[\rho_2^*]$. Define the *normalized harmonic mean share weights* as $s_n^* \equiv h(s_{1n},s_{2n})/\Sigma_{i=1}^N h(s_{1i},s_{2i})$ for n = 1,...,N. Then the weighted time product dummy bilateral price index, $P_{WTPD}(p^1,p^2,q^1,q^2) \equiv \pi_2^*/\pi_1^* = \pi_2^*$, has the following logarithm:

$$(64)\ \ln P_{WTPD}(p^1,p^2,q^1,q^2) \equiv \Sigma_{n=1}^N s_n^* \ln(p_{2n}/p_{1n}).$$

Thus $P_{WTPD}(p^1,p^2,q^1,q^2)$ is equal to a share weighted geometric mean of the price ratios, p_{2n}/p_{1n}.[66] This index is a satisfactory one from the viewpoint of the test approach to index number theory. It can be shown that $P_{WTPD}(p^1,p^2,q^1,q^2)$ satisfies the following tests: (i) the identity test; i.e., $P_{WTPD}(p^1,p^2,q^1,q^2) = 1$ if $p^1 = p^2$; (ii) the time reversal test; i.e., $P_{WTPD}(p^2,p^1,q^2,q^1) = 1/P_{WTPD}(p^1,p^2,q^1,q^2)$;[67] (iii) homogeneity of degree 1 in period 2 prices; i.e., $P_{WTPD}(p^1,\lambda p^2,q^1,q^2) = \lambda P_{WTPD}(p^1,p^2,q^1,q^2)$; (iv) homogeneity of degree −1 in period 1 prices; i.e., $P_{WTPD}(\lambda p^1,p^2,q^1,q^2) = \lambda^{-1}P_{WTPD}(p^1,p^2,q^1,q^2)$; (v) homogeneity of degree 0 in period 1 quantities; i.e., $P_{WTPD}(p^1,p^2,\lambda q^1,q^2) = \lambda P_{WTPD}(p^1,p^2,q^1,q^2)$; (vi) homogeneity of degree 0 in period 2 quantities; i.e., $P_{WTPD}(p^1,p^2,q^1,\lambda q^2) =$

[65] Note that the minimization problem defined by (61) is equivalent to the problem of minimizing $\Sigma_{n=1}^N e_{1n}^2 + \Sigma_{n=1}^N e_{2n}^2$ with respect to $\rho_2, \beta_1, ..., \beta_N$ where the error terms e_{tn} are defined by the equations $s_{1n}^{1/2}\ln p_{1n} = s_{1n}^{1/2}\beta_n + e_{1n}$ for n = 1,...,N and $s_{2n}^{1/2}\ln p_{2n} = s_{2n}^{1/2}\rho_2 + s_{2n}^{1/2}\beta_n + e_{2n}$ for n = 1,...,N. Thus the solution to (61) can be found by running a linear regression using the above two sets of estimating equations. The numerical equivalence of the least squares estimates obtained by repeating multiple observations or by using 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).

[66] See Diewert (2002) (2005a).

[67] See Diewert (2003b) (2005b).

$P_{WTPD}(p^1,p^2,q^1,q^2)$; (vii) invariance to changes in the units of measurement; (viii) $\min_n\{p_{2n}/p_{1n} : n = 1,...,N\} \leq P_{WTPD}(p^1,p^2,q^1,q^2) \leq \max_n\{p_{2n}/p_{1n} : n = 1,...,N\}$; and (ix) invariance to the ordering of the products. Moreover, Diewert (2005b; 564) showed that $P_{WTPD}(p^1,p^2,q^1,q^2)$ approximated the superlative Törnqvist-Theil index to the second order around an equal price and quantity point where $p^1 = p^2$ and $q^1 = q^2$.[68] Thus if changes in prices and quantities going from one period to the next are not too large and there are no missing products, P_{WTPD} should be close to the superlative Fisher (1922) and Törnqvist-Theil indexes.[69]

Recall the results from section 5 above for the unweighted time product dummy model. From equation (38), it can be seen that the unweighted bilateral time product dummy regression model generated the Jevons index as the solution to the unweighted least squares minimization problem that is a counterpart to the weighted problem defined by (61) above. Thus appropriate weighting of the squared errors has changed the solution index dramatically: the index defined by (64) weights products by their economic importance and has good test properties whereas the Jevons index can generate very problematic results due to its lack of weighting according to economic importance. Note that both models have the same underlying structure; i.e., they assume that p_{tn} is approximately equal to $\pi_t\alpha_n$ for t = 1,2 and n = 1,...,N. *Thus weighting by economic importance has converted a least squares minimization problem that generates a rather poor price index into a problem that generates a rather good index.*

There is one more weighting scheme that generates an even better index in the bilateral context where we are running a time product dummy hedonic regression using the price and quantity data for only two periods. Consider the following weighted least squares minimization problem:

$$(65)\ \min_{\rho,\beta}\ \{\Sigma_{n=1}^{N} (½)(s_{1n}+s_{2n})[\ln p_{1n} - \beta_n]^2 + \Sigma_{n=1}^{N} (½)(s_{1n}+s_{2n})[\ln p_{2n} - \rho_2 - \beta_n]^2\}.$$

As usual, it can be seen that problem (65) has exactly the same mathematical form as problem (51) except that $(½)(s_{1n}+s_{2n})$ has replaced q_{tn} and so the solutions (54) and (55) will be valid in the present context if $(½)(s_{1n}+s_{2n})$ replaces q_{tn} in these formulae. Thus the solution to (65) simplifies to the following solution:

$$(66)\ \rho_2^* \equiv \Sigma_{n=1}^{N} (½)(s_{1n}+s_{2n})\ln(p_{2n}/p_{1n});$$
$$(67)\ \beta_n^* \equiv (½)\ln(p_{1n}) + (½)\ln(p_{2n}/\pi_2^*)\ ; \qquad n = 1,...,N$$

where $\pi_2^* \equiv \exp[\rho_2^*]$ and $\pi_1^* \equiv \exp[\rho_1^*] = \exp[0] = 1$ since we have set $\rho_1^* = 0$. Thus the bilateral index number formula which emerges from the solution to (65) is π_2^*/π_1^* =

[68] Diewert (2005a; 564) noted this result. Thus P_{WTPD} is a pseudo-superlative index. For the definition of a superlative index, see Diewert (1976). A pseudo-superlative index approximates a superlative index to the second order around any point where $p^1 = p^2$ and $q^1 = q^2$; see Diewert (1978).
[69] However, with large changes in price and quantities going from period 1 to 2, P_{WTPD} will tend to lie below its superlative counterparts; see Diewert (2018; 53) and the example in Diewert and Fox (2017; 24).

$\exp[\Sigma_{n=1}^{N} (½)(s_{1n}+s_{2n})\ln(p_{2n}/p_{1n})] \equiv P_T(p^1,p^2,q^1,q^2)$, which is the Törnqvist (1936)[70] Theil (1967; 137-138) bilateral index number formula.[71] Thus the use of the weights in (65) has generated an even better bilateral index number formula than the formula which resulted from the use of the weights in (61).[72] This result reinforces the case for using appropriately weighted versions of the basic time product dummy hedonic regression model.[73] However, if the implied residuals in the minimization problem (65) are small (or equivalently, if the fit in the linear regression model that can be associated with (65) is high so that predicted values for log prices are close to actual log prices), then *weighting will not matter very much* and the unweighted version of (65) will give results that are similar to (65). This comment applies to all of the weighted hedonic regression models that are considered in this paper.[74]

The aggregate quantity levels for the t periods can be obtained as $Q^{t*} \equiv \alpha^*\cdot q^t = \Sigma_{n=1}^{N} \alpha_n^* q_{tn}$ for t = 1,2 where the α_n^* are defined as the exponentials of the β_n^* defined by (67). Estimated aggregate price levels can be obtained directly from the solution to (65); i.e., set $P^{t*} = \pi_t^*$ for t = 1,2.[75] Alternative price levels can be obtained indirectly as $P^{t**} \equiv p^t\cdot q^t/Q^{t*} = p^t\cdot q^t/\alpha^*\cdot q^t$ for t = 1,2. If the optimized objective function in (65) is 0, so that all errors equal 0, then P^{t*} will equal P^{t**} for t = 1,2. If the estimated residuals are not all equal to 0, then the two estimates for the period t price level P^t will differ and the alternative estimates for P^t will generate different estimates for the companion aggregate quantity levels.

It should be noted that we have not made any bias corrections due to the fact that our model estimates the logarithm of π_t instead of π_t itself. This is due to our perspective that simply tries to fit an exact model by transforming it in a way that leads to solutions π_t^* to a least squares minimization problem where the π_t^* have good axiomatic properties.[76]

---

[70] See Törnqvist and Törnqvist (1937) for the actual formula. Diewert (1976) showed that this index number formula was superlative.

[71] An alternative to Theil's (1967; 137-138) derivation of $P_T(p^1,p^2,q^1,q^2)$ can be obtained by solving the following least squares minimization problem: $\min_\rho \Sigma_{n=1}^{N} (½)(s_{1n}+s_{2n})[\ln(p_{2n}/p_{1n}) - \rho]^2$ where $\rho \equiv \ln P_T(p^1,p^2,q^1,q^2)$. The β_n do not appear in this minimization problem.

[72] Diewert (1992; 223) lists the commonly used tests that $P_T(p^1,p^2,q^1,q^2)$ satisfies.

[73] Note that the bilateral regression model defined by the minimization problem (61) is readily generalized to the case of T periods whereas the bilateral regression model defined by the minimization problem (65) cannot be generalized to the case of T periods. These facts were noted by de Haan and Krsinich (2012).

[74] If the residuals are small for (65), then prices will vary almost proportionally over time and all reasonable index number formulae will register price levels that are close to the estimated π_t^*; i.e., we will have $p^t \approx \pi_t^* p^1$ for t = 2,3,...,T with $\pi_1^* = 1$.

[75] In this case, alternative period t quantity levels are defined as $Q^{1**} \equiv p^1\cdot q^1$ and $Q^{2**} \equiv p^2\cdot q^2/\pi_2^* = [v_2/v_1]/P_T(p^1,p^2,q^1,q^2)$. If the squared errors in (65) are all 0, then the alternative quantity estimates are equal to each other and the model $\ln p_{tn} = \rho_t + \beta_n$ holds exactly for each tn and prices are proportional across the two periods; i.e., we have $p^t = \pi_t^*\alpha^*$ for t = 1,2 where $\alpha^* \equiv [\alpha_1^*,...,\alpha_N^*]$. In the case where the squared errors are nonzero, the π_t^*,Q^{t**} aggregates are preferred since $P_T(p^1,p^2,q^1,q^2)$ is a superlative index and thus has a strong economic justification.

[76] We note that de Haan and Krsinich (2018; 769-770) make the following comments on possible biases that result from the use of a weighted least squares model to generate price indexes: "Finally, we will elaborate on a few econometric issues. The estimated quality adjusted prices ... are biased as taking exponentials is a non-linear transformation. The time dummy index is similarly biased. It is questionable

There is more work to be done in working out the distributional properties of the above estimators for the price levels.

8. Weighted Time Product Dummy Regressions: The Bilateral Case with Missing Observations

In this section, we will generalize the last two models in the previous section to cover the case where there are missing observations.[77] Thus we assume that there are products that are missing in period 2 that were present in period 1 and some new products that appear in period 2. As in section 6 above, S(t) denotes the set of products n that are present in period t for t = 1,2. It is assumed that $S(1)\cap S(2)$ is not the empty set; i.e., there are one or more products that are present in both periods. The new weighted least squares minimization problem that generalizes (61) is the following minimization problem:[78]

$$(68)\ \min_{\rho,\beta} \sum_{n\in S(1)} s_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n\in S(2)} s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2.$$

The first order conditions for ρ_2^*, β_1^*,..., β_N^* to solve (68) are equivalent to the following equations:

$$(69)\ \sum_{n\in S(2)} s_{2n}[\ln p_{2n} - \rho_2^* - \beta_n^*] = 0\ ;$$
$$(70)\ (s_{1n} + s_{2n})\beta_n^* = s_{1n}\ln p_{1n} + s_{2n}[\ln p_{1n} - \rho_2^*]\ ; \qquad n\in S(1)\cap S(2);$$
$$(71)\qquad \beta_n^* = \ln p_{1n}\ ; \qquad n\in S(1),\ n\notin S(2);$$
$$(72)\qquad \beta_n^* = \ln p_{2n} - \rho_2^*\ ; \qquad n\in S(2),\ n\notin S(1).$$

Define the intersection set of products S^* as follows:

$$(73)\ S^* \equiv S(1)\cap S(2).$$

Substituting equations (72) into equation (69) leads to the following equation:

$$(74)\ \sum_{n\in S^*} s_{2n}[\ln p_{2n} - \rho_2^* - \beta_n^*] = 0.$$

Consider the following least squares minimization problem that is defined over the set of products that are present in both periods:

$$(75)\ \min_{\rho,\beta,n\in S^*} \sum_{n\in S^*} s_{1n}[\ln p_{1n} - \beta_n]^2 + \sum_{n\in S^*} s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2.$$

whether bias adjustments would be appropriate, though, at least from an index number point of view. For instance, recall the two-period case with only matched items, where Diewert's (2004) choice of regression weights ensures that the time dummy index is equal to the superlative Törnqvist price index. Correcting for the "bias" would mean that this useful property does no longer hold, and so there is a tension between econometrics and index number theory."

[77] The results in this section are closely related to the results derived by de Haan (2004a), Silver and Heravi (2005) and de Haan and Krsinich (2014) (2018). However, our method of derivation is somewhat different.

[78] This form of weighting was suggested by Rao (1995) (2004) (2005), Diewert (2002) (2004) (2005a) and de Haan (2004a).

The first order conditions for this problem are (74) and (70). Once we find the solution to this problem, define β_n^* for the products that are not present in both periods by equations (71) and (72). This augmented solution will solve problem (68). The solution to (75) can be found by adapting the solution to (61) to the current situation. Recall equations (62) and (63) from the previous section. Replacing the entire set of product indices n = 1,...,N by the intersection set S^* defined by (73) leads to the following solution to (75):

$$(76)\ \rho_2^* \equiv [\Sigma_{n\in S^*}\ s_{1n}s_{2n}(s_{1n} + s_{2n})^{-1} \ln(p_{2n}/p_{1n})]/[\Sigma_{i\in S^*}\ s_{1i}s_{2i}(s_{1i} + s_{2i})^{-1}] ;$$

$$(77)\ \beta_n^* \equiv s_{1n}(s_{1n} + s_{2n})^{-1} \ln(p_{1n}) + s_{2n}(s_{1n} + s_{2n})^{-1} \ln(p_{2n}/\pi_2^*) ; \qquad n\in S^*$$

where $\pi_2^* \equiv \exp[\rho_2^*]$. Define the *normalized harmonic mean share weights* for the always present products as follows:

$$(78)\ s_n^* \equiv h(s_{1n},s_{2n})/\Sigma_{i\in S^*} h(s_{1i},s_{2i}); \qquad n\in S^*.$$

Denote the period t price and quantity vectors that include only matched products by p^{t*} and q^{t*} respectively for t = 1,2. Then the *weighted time product dummy bilateral price index with missing observations*, $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \pi_2^*/\pi_1^* = \pi_2^*$, has the following logarithm:

$$(78)\ \ln P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \Sigma_{n\in S^*}\ s_n^* \ln(p_{2n}/p_{1n}).$$

Note that $P_{WTPD} \equiv \pi_2^*/\pi_1^*$ depends only on the price and share information for the products that are present in both periods. Thus the bilateral price index that is generated by solving (68) is similar to the nonstochastic maximum overlap price index that was defined in section 4. The main difference is that the maximum overlap Fisher price index P_F that was used in section 4 is superlative whereas the present index, P_{WTPD}, is only pseudo-superlative. Thus $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*})$ approximates $P_F(p^{1*},p^{2*},q^{1*},q^{2*})$ to the second order around any point where $p^{1*} = p^{2*}$ and $q^{1*} = q^{2*}$. However, if the products under consideration are subject to frequent price discounts, the fluctuations in prices and quantities can be huge and second order approximations may not be very accurate. Under these circumstances, it seems likely that the section 4 option is the better choice for statistical agencies; it is simpler to explain to the public and the axiomatic properties of the Fisher index are likely to be better than the axiomatic properties of P_{WTPD}.

To complete the solution to (68), use equations (71) and (72) to define the β_n^* for the new and missing products in period 2. Thus P_{WTPD} *uses only the matched prices* to define the price index that results from solving the weighted least squares minimization problem with missing observations that is defined by (68).

As usual, the hedonic regression model that is generated by solving (68) can be used to impute reservation prices for missing observations. Thus define $\alpha_n^* \equiv \exp[\beta_n^*]$ for n = 1,...,N. Then the missing prices p_{tn}^* can be defined as follows:

$$(79)\ p_{2n}^* \equiv \pi_2^*\alpha_n^* = \pi_2^* p_{1n} \qquad n\in S(1),\ n\notin S(2);$$

$$(80)\ p_{1n}^* \equiv \pi_1^*\alpha_n^* = p_{2n}/\pi_2^* \qquad n\in S(2),\ n\notin S(1).$$

Thus the missing prices for period 2, p_{2n}^*, are the corresponding *inflation adjusted carry forward prices* from period 1, p_{1n} times π_2^* and the missing prices for period 1, p_{1n}^*, are the corresponding *inflation adjusted carry backward prices* from period 2, p_{2n} deflated by π_2^*, where π_2^* is the weighted time product dummy price index $P_{WTPDM}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (78). Note the similarity of equations (79) and (80) to the nonstochastic imputed prices defined by equations (25) in section 4.

Estimated aggregate price levels can be obtained directly from the solution to (75); i.e., set $P^{1*} = 1$ and $P^{2*} = \pi_2^*$. The corresponding quantity levels are defined as $Q^{1*} \equiv p^1 \cdot q^1$ and $Q^{2*} \equiv p^2 \cdot q^2/\pi_2^*$.[79] Alternative price and quantity levels can be obtained as $Q^{t**} \equiv \alpha^* \cdot q^t$ and $P^{t**} \equiv p^t \cdot q^t/Q^{t**}$ for t = 1,2. If the optimized objective function in (75) is 0, so that all errors equal 0, then P^{t*} will equal P^{t**} for all t. If the estimated residuals are not all equal to 0, then the two estimates for the period t price level P^t will differ and, as usual, the alternative estimates for P^t will generate different estimates for the companion aggregate quantity levels.

The above analysis is not quite the end of the story. The expenditure shares s_{1n} and s_{2n} which appear in (75) are not the expenditure shares that characterize the always present products; they are the original expenditure shares defined over all N products. It is of interest to compare $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (78) with the weighted time product dummy index, $P_{WTPDM}(p^{1*},p^{2*},q^{1*},q^{2*})$, that is defined over the common set of products, S^*;[80] i.e., P_{WTPDM} is the weighted time product dummy regression model that is defined over the set of *matched products* for the two periods under consideration.

Define $v_t^* \equiv \Sigma_{n\in S^*} v_{tn}$ as the total expenditure on always present products for t = 1,2 and define the corresponding *restricted expenditure shares* as:

$$(81)\quad s_{tn}^* \equiv v_{tn}/v_t^* ; \qquad\qquad t = 1,2;\ n\in S^*.$$

The matched model version of (75) is the following weighted least squares minimization problem:

$$(82)\quad \min_{\rho,\, \beta,\, n\in S^*} \Sigma_{n\in S^*}\ s_{1n}^*[\ln p_{1n} - \beta_n]^2 + \Sigma_{n\in S^*}\ s_{2n}^*[\ln p_{2n} - \rho_2 - \beta_n]^2.$$

The ρ_2 solution to (82) is the following one:

$$\begin{aligned}(83)\quad \rho_2^{**} &\equiv [\Sigma_{n\in S^*}\ s_{1n}^* s_{2n}^* (s_{1n}^* + s_{2n}^*)^{-1} \ln(p_{2n}/p_{1n})]/[\Sigma_{i\in S^*}\ s_{1i}^* s_{2i}^* (s_{1i}^* + s_{2i}^*)^{-1}] \\ &= [\Sigma_{n\in S^*}\ h(s_{1n}^*,s_{2n}^*) \ln(p_{2n}/p_{1n})]/[\Sigma_{i\in S^*}\ h(s_{1i}^*,s_{2i}^*)]\end{aligned}$$

---

[79] Use the imputed prices defined by (78) and (79) for the missing prices and set the missing quantity levels equal to 0. The resulting N dimensional vectors are denoted by p^t and q^t for t = 1,2 in the definitions which follow.

[80] Recall that S^* is defined by (73) and p^{t*} and q^{t*} are the period t price and quantity vectors that include only products that are present in both periods.

where $h(s_{1n}^*,s_{2n}^*)$ is the harmonic mean of the restricted shares s_{1n}^* and s_{2n}^*. Thus $P_{WTPDM}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \exp[\rho_2^{**}]$ where ρ_2^{**} is defined by (83).

The relationship between the *true shares*, the s_{tn}, and the *restricted shares*, the s_{tn}^*, for the always present products is given by the following equations:

$$(84)\ s_{tn} \equiv v_{tn}/v_t = [v_{tn}/v_t^*][v_t^*/v_t] = s_{tn}^* f_t \ ; \qquad t = 1,2 \ ; n \in S^*$$

where the *fraction* of expenditures on always available commodities compared to expenditures on all commodities during period t is $f_t \equiv v_t^*/v_t$ for t = 1,2. Using equations (84) and (76), it can be seen that the logarithm of $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (78) is equal to the following expression:

$$\begin{aligned}(85)\ \rho_2^* &\equiv [\Sigma_{n\in S^*}\ h(s_{1n},s_{2n}) \ln(p_{2n}/p_{1n})]/[\Sigma_{i\in S^*}\ h(s_{1i},s_{2i})] \\ &= [\Sigma_{n\in S^*}\ h(f_1 s_{1n}^*,f_2 s_{2n}^*) \ln(p_{2n}/p_{1n})]/[\Sigma_{i\in S^*}\ h(f_1 s_{1i}^*,f_2 s_{2i}^*)].\end{aligned}$$

Now compare (83) and (85). If either: (i) $p_{2n} = \lambda p_{1n}$ for all $n \in S^*$ so that we have price proportionality for the always present products or (ii) $f_1 = f_2$ so that the ratio of expenditures on always present products to total expenditure in each period is constant across the two periods, then $\rho_2^{**} = \rho_2^*$. However, if these conditions are not satisfied and there is considerable variation in prices and quantities across periods, then ρ_2^{**} could differ substantially from ρ_2^*. Since neither index is superlative, it is difficult to recommend one of these indexes over the other as the "optimal" carry forward and backward inflation rate which could be used to construct the inflation adjusted carry forward and backward estimates for the missing prices.

We conclude this section by considering the weighting scheme suggested by de Haan (2004a). The following weighted least squares minimization problem is his generalization of (65) to the case of missing observations:[81]

$$\begin{aligned}(86)\ \min_{\rho,\beta}\ & \Sigma_{n\in S^*} (½)(s_{1n}+s_{2n})[\ln p_{1n} - \beta_n]^2 + \Sigma_{n\in S(1),\, n\notin S(2)} (½)s_{1n}[\ln p_{1n} - \beta_n]^2 \\ &+ \Sigma_{n\in S^*} (½)(s_{1n}+s_{2n})[\ln p_{2n} - \rho_2 - \beta_n]^2 + \Sigma_{n\in S(2),\, n\notin S(1)} (½)s_{2n}[\ln p_{2n} - \rho_2 - \beta_n]^2.\end{aligned}$$

The first order conditions that a solution to (86) satisfies simplify to the following conditions:

$$\begin{aligned}(87)\ & \Sigma_{n\in S^*} (s_{1n}+s_{2n})\rho_2^* + \Sigma_{n\in S(2),\, n\notin S(1)}\ s_{2n}\rho_2^* \\ &\quad = \Sigma_{n\in S^*} (s_{1n}+s_{2n})[\ln p_{2n} - \beta_n^*] + \Sigma_{n\in S(2),\, n\notin S(1)}\ s_{2n}[\ln p_{2n} - \beta_n^*] \ ;\end{aligned}$$

$$(88)\ 2\beta_n^* = \ln p_{1n} + \ln p_{2n} - \rho_2^* \ ; \qquad n\in S^*;$$

$$(89)\ \ \beta_n^* = \ln p_{1n} \ ; \qquad n\in S(1),\ n\notin S(2);$$

$$(90)\ \ \beta_n^* = \ln p_{2n} - \rho_2^* \ ; \qquad n\in S(2),\ n\notin S(1).$$

[81] Recall that the weighting scheme used in (65) generated the "best" index number formula in the previous section where there were no missing observations.

Substitute equations (88) and (90) into equation (87). The resulting equation simplifies to the following equation:

$$(91)\ \rho_2^* = \Sigma_{n\in S^*} \tfrac{1}{2}(s_{1n}+s_{2n})\ln(p_{2n}/p_{1n})/\Sigma_{i\in S^*} \tfrac{1}{2}(s_{1i}+s_{2i})$$
$$= \Sigma_{n\in S^*} \tfrac{1}{2}(f_1 s_{1n}^*+f_2 s_{2n}^*)\ln(p_{2n}/p_{1n})/\Sigma_{i\in S^*} \tfrac{1}{2}(f_1 s_{1i}^*+f_2 s_{2i}^*)$$

where the second equation follows using equations (84). Recall that the s_{tn} are the true expenditure shares and the s_{tn}^* are restricted expenditure shares that sum to one when only always present products are in scope. Once ρ_2^* has been determined, the β_n^* are determined using equations (88)-(90). As usual, define $\pi_1^* \equiv 1$ and $\pi_2^* \equiv \exp[\rho_2^*]$ where ρ_2^* is defined by (91). Then the *Haan weighted time product dummy bilateral price index with missing observations*, $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \pi_2^*/\pi_1^* = \pi_2^*$, has the following logarithm:

$$(92)\ \ln P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \Sigma_{n\in S^*}\ s_n^* \ln(p_{2n}/p_{1n})$$

where the shares s_n^* are defined as follows:

$$(93)\ s_n^* \equiv \tfrac{1}{2}(f_1 s_{1n}^*+f_2 s_{2n}^*)/\Sigma_{i\in S^*} \tfrac{1}{2}(f_1 s_{1i}^*+f_2 s_{2i}^*)\ ; \qquad n\in S^*.$$

If either: (i) $p_{2n} = \lambda p_{1n}$ for all $n\in S^*$ so that we have price proportionality for the always present products or (ii) $f_1 = f_2$ so that the ratio of expenditures on always present products to total expenditure in each period is constant across the two periods, then $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*}) = P_T(p^{1*},p^{2*},q^{1*},q^{2*})$, the Törnqvist Theil index defined over the always present products.[82] Note that P_{WTPDH} *uses only the matched prices* to define the price index that results from solving the weighted least squares minimization problem with missing observations that is defined by (86).

As usual, the bilateral index $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*}) \equiv \exp[\rho_2^*] \equiv \pi_2^*$ where ρ_2^* is defined by (91) can be used to impute reservation prices for missing observations. Thus define $\alpha_n^* \equiv \exp[\beta_n^*]$ for n = 1,...,N. Then the missing prices p_{tn}^* can be defined as follows:

$$(94)\ p_{2n}^* \equiv \pi_2^*\alpha_n^* = \pi_2^* p_{1n} \qquad n\in S(1),\ n\notin S(2);$$
$$(95)\ p_{1n}^* \equiv \pi_1^*\alpha_n^* = p_{2n}/\pi_2^* \qquad n\in S(2),\ n\notin S(1).$$

Thus the missing prices for period 2, the p_{2n}^* defined by (94), are the corresponding *inflation adjusted carry forward prices* from period 1, p_{1n} times π_2^* and the missing prices for period 1, p_{1n}^* defined by (95), are the corresponding *inflation adjusted carry backward prices* from period 2, p_{2n} deflated by π_2^*, where π_2^* is the de Haan weighted time product dummy price index $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (92). Again, note the similarity of equations (94) and (95) to the nonstochastic imputed prices defined by equations (25) in section 4.

[82] This result was first derived by de Haan and Krsinich (2014; 347).

Comparing $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (78) and $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (92), the latter index is preferred due to the fact that it collapses to the superlative Törnqvist Theil index $P_T(p^{1*},p^{2*},q^{1*},q^{2*})$ if either there are no missing observations or if $f_1 = f_2$. If $f_1 = f_2$, $P_{WTPD}(p^{1*},p^{2*},q^{1*},q^{2*})$ does not collapse to $P_T(p^{1*},p^{2*},q^{1*},q^{2*})$.

Finally, de Haan and Krsinich (2014; 345) derived an interesting property of the Haan index defined by (92): define the missing shares and quantities as follows:

(96) $s_{2n} \equiv 0$ and $q_{2n} \equiv 0$; $\quad n \in S(1),\ n \notin S(2)$;
(97) $s_{1n} \equiv 0$ and $q_{2n} \equiv 0$; $\quad n \in S(2),\ n \notin S(1)$.

Define the missing prices by (94) and (95). Denote the now complete price and quantity vectors for period t as p^t and q^t for t = 1,2. Calculate the Törnqvist Theil index over the complete data as $P_T(p^1,p^2,q^1,q^2)$. It turns out that $P_T(p^1,p^2,q^1,q^2) = P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*})$ where the logarithm of the latter index is defined by (91).

To prove this result, note that for $n \in S(2)$, $n \notin S(1)$, we have $s_{1n} = 0$ using (97). Thus we can add $\Sigma_{n\in S(2),n\notin S(1)}\, s_{1n}\rho_2^*$ to the left hand side of (87) and add $\Sigma_{n\in S(2),n\notin S(1)}\, s_{1n}[\ln p_{2n} - \beta_n^*]$ to the right hand side of (87). The resulting equation is:

$$(98)\ \Sigma_{n\in S^*}(s_{1n}+s_{2n})\rho_2^* + \Sigma_{n\in S(2),\, n\notin S(1)}\,(s_{1n} + s_{2n})\rho_2^* \\ = \Sigma_{n\in S^*}(s_{1n}+s_{2n})[\ln p_{2n} - \beta_n^*] + \Sigma_{n\in S(2),\, n\notin S(1)}\,(s_{1n}+s_{2n})[\ln p_{2n} - \beta_n^*]\ .$$

Upon taking logarithms of both sides of equations (94), it can be seen that we can add $\Sigma_{n\in S(1),n\notin S(2)}\,(s_{1n}+s_{2n})\rho_2^*$ to the left hand side of (87) and add $\Sigma_{n\in S(2),n\notin S(1)}\,(s_{1n}+s_{2n})[\ln p_{2n}^* - \beta_n^*]$ to the right hand side of (98). The resulting equation is:

$$(99)\ \Sigma_{n=1}^{N}(s_{1n}+s_{2n})\rho_2^* = \Sigma_{n=1}^{N}(s_{1n}+s_{2n})[\ln p_{2n} - \beta_n^*]$$

where p_{2n}^* is denoted as p_{2n} for $n \in S(1)$ and $n \notin S(2)$ in (99). Our next task is to eliminate the sum $\Sigma_{n=1}^{N}(s_{1n}+s_{2n})\beta_n^*$ from the right hand side of (99). Using equations (88), it is straightforward to derive the following equation:

$$(100)\ \Sigma_{n\in S^*}(s_{1n}+s_{2n})\beta_n^* = ½\Sigma_{n\in S^*}(s_{1n}+s_{2n})[\ln p_{1n} + \ln p_{2n} - \rho_2^*].$$

Upon taking logarithms of both sides of equations (94), we obtain the equations $\beta_n^* = \ln p_{2n}^* - \rho_2^*$ for $n \in S(1)$, $n \notin S(2)$. Using these equations and equations (89), it can be seen that we obtain the equations $2\beta_n^* = [\ln p_{1n} + \ln p_{2n}^* - \rho_2^*]$ for $n \in S(1)$, $n \notin S(2)$ and these equations imply the following equation:

$$(101)\ \Sigma_{n\in S(1),n\notin S(2)}(s_{1n}+s_{2n})\beta_n^* = ½\Sigma_{n\in S(1),n\notin S(2)}(s_{1n}+s_{2n})[\ln p_{1n} + \ln p_{2n}^* - \rho_2^*].$$

Upon taking logarithms of both sides of equations (95), we obtain the equations $\beta_n^* = \ln p_{1n}^*$ for $n \in S(2)$, $n \notin S(1)$. Using these equations and equations (90), it can be seen that we obtain the equations $2\beta_n^* = [\ln p_{1n}^* + \ln p_{2n} - \rho_2^*]$ for $n \in S(2)$, $n \notin S(1)$ and these equations imply the following equation:

(102) $\Sigma_{n\in S(2),n\notin S(1)} (s_{1n}+s_{2n})\beta_n^* = ½\Sigma_{n\in S(2),n\notin S(1)} (s_{1n}+s_{2n})[\ln p_{1n}^* + \ln p_{2n} - \rho_2^*]$.

Equations (100)-(102) imply that the following equation holds:

(103) $\Sigma_{n=1}^N (s_{1n}+s_{2n})\beta_n^* = ½\Sigma_{n=1}^N (s_{1n}+s_{2n})[\ln p_{1n} + \ln p_{2n} - \rho_2^*]$

where we have denoted the implicit prices p_{tn}^* in (103) by p_{tn} where necessary. Finally substitute (103) into (99) and using $\Sigma_{n=1}^N s_{tn} = 1$ for $t = 1,2$, we obtain the following expression for ρ_2^*:

(104) $\rho_2^* = ½\Sigma_{n=1}^N (s_{1n}+s_{2n})\ln(p_{2n}/p_{1n})$.

Thus the ρ_2^* defined by (91) is also equal to the ρ_2^* defined by (104) and thus the Haan index $P_{WTPDH}(p^{1*},p^{2*},q^{1*},q^{2*})$ defined by (92) over the set of products that are available in both periods is also equal to the Törnqvist Theil index over the complete data, $P_T(p^1,p^2,q^1,q^2)$, where the missing prices are defined by (94) and (95).

Our conclusion at this point is that the Haan system of weighting when there are missing observations and only two periods is the preferred method of weighting.[83]

In the following section, we define weighted time dummy regression models for the general case of T periods.

9. Weighted Time Product Dummy Regressions: The General Case

We first consider the case of no missing observations. The generalization of the two period weighted least squares minimization problem that was defined by (61) in section 7 to the case of $T > 2$ periods is (105) below:[84]

(105) $\min_{\rho,\beta} \Sigma_{n=1}^N \Sigma_{t=1}^T s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2$.

The first order necessary conditions for $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ to solve (105) are the following T equations (106) and N equations (107):

(106) $\rho_t^* = \Sigma_{n=1}^N s_{tn}[\ln p_{tn}^* - \beta_n^*]$; $t = 1,...,T$;
(107) $\beta_n^* = \Sigma_{t=1}^T s_{tn}[\ln p_{tn}^* - \rho_t^*]/(\Sigma_{t=1}^T s_{tn})$; $n = 1,...,N$.

---

[83] However, the fact that the Haan index depends only on the matched prices and quantities means and that the missing prices are inflation adjusted carry forward and backward prices means that the Haan index is simply a substitute for other possible indexes defined over the set of matched products; i.e., it is not clear that the Haan index has better properties than the index used in section 4 to generate inflation adjusted carry forward and backward prices.
[84] Rao (1995) (2004) (2005; 574) was the first to consider this model using expenditure share weights. However, Balk (1980; 70) suggested this class of models much earlier using somewhat different weights.

As usual, the solution to (106) and (107) is not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (106) and (107), then so do $[\rho_1^*+\lambda,...,\rho_T^*+\lambda]$ and $[\beta_1^*-\lambda,...,\beta_N^*-\lambda]$ for all λ. Thus we can set $\rho_1^* = 0$ in equations (107) and drop the first equation in (106) and use linear algebra to find a unique solution for the resulting equations.[85] Once the solution is found, define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* as follows:

(108) $\pi_t^* \equiv \exp[\rho_t^*]$; $t = 1,...,T$; $\alpha_n^* \equiv \exp[\beta_n^*]$; $n = 1,...,N$.

Note that the resulting *price index* between periods t and r is equal to the following expression:

(109) $\pi_t^*/\pi_r^* = \prod_{n=1}^N \exp[s_{tn}\ln(p_{tn}/\alpha_n^*)]/\prod_{n=1}^N \exp[s_{rn}\ln(p_{rn}/\alpha_n^*)]$; $\qquad 1 \leq t, \tau \leq T$.

If $s_{tn} = s_{rn}$ for $n = 1,...,N$, then π_t^*/π_r^* will equal a weighted geometric mean of the price ratios $p_{tn}/p_{\tau n}$ where the weight for $p_{tn}/p_{\tau n}$ is the common expenditure share $s_{tn} = s_{\tau n}$. Thus π_t^*/π_τ^* will not depend on the α_n^* in this case.

The price levels π_t^* defined by (108) are functions of the T price vectors, $p^1,...,p^T$ and the t quantity vectors $q^1,...,q^T$. These price level functions have some good axiomatic properties: (i) the π_t^* are invariant to changes in the units of measurement; (ii) π_t^* regarded as a function of the period t price vector p^t is linearly homogeneous in the components of p^t; i.e., $\pi_t^*(\lambda p^t) = \lambda\pi_t^*(p^t)$ for all $p^t >> 0_N$ and $\lambda > 0$; (iii) π_t^* regarded as a function of the period t quantity vector q^t is homogeneous of degree 0 in the components of q^t; i.e., $\pi_t^*(\lambda q^t) = \pi_t^*(q^t)$ for all $q^t >> 0_N$ and $\lambda > 0$;[86] (iv) the π_t^* satisfy a version of Walsh's (1901; 389) (1921b; 540) *multiperiod identity test*; i.e., if $p^t = p^\tau$ and $q^t = q^\tau$, then $\pi_t^* = \pi_\tau^*$.[87]

Once the estimates for the π_t and α_n have been computed, we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for $t = 1,...,T$. The π_t^* estimates can be used to form the aggregates using equations (110) or the α_n^* estimates can be used to form the aggregates using equations (111):[88]

(110) $P^{t*} \equiv \pi_t^*$; $\qquad Q^{t*} \equiv p^t \cdot q^t/\pi_t^*$; $\qquad t = 1,...,T$;

---

[85] Alternatively, one can set up the linear regression model defined by $(s_{tn})^{1/2}\ln p_{tn} = (s_{tn})^{1/2}\rho_t + (s_{tn})^{1/2}\beta_n + e_{tn}$ for $t = 1,...,T$ and $n = 1,...,N$ where we set $\rho_1 = 0$ to avoid exact multicollinearity. Iterating between equations (106) and (107) will also generate a solution to these equations and the solution can be normalized so that $\rho_1 = 0$.

[86] By looking at the minimization problem defined by (105), it is also straightforward to show that $\pi_t^*(\lambda q^\tau) = \pi_t^*(q^\tau)$ for all $q^\tau >> 0_N$ and $\lambda > 0$ for $\tau = 1,....,T$.

[87] We would like the π_t^* to satisfy the usual (strong) identity test which is: if $p^t = p^\tau$, then $\pi_t^* = \pi_\tau^*$. However, if the share weights for the two periods are different, then this test no longer holds. However, if we define the period t price and quantity levels using definitions (111), it can be seen that the resulting Q^{t**} will satisfy the usual (strong) identity test for quantities. This is a good reason for choosing (111) over (110).

[88] Note that the price level P^{t**} defined in (111) is a quality adjusted unit value index of the type studied by de Haan (2004b).

(111) $Q^{t**} \equiv \alpha^*\cdot q^t$; $P^{t**} \equiv p^t\cdot q^t/\alpha^*\cdot q^t$; $t = 1,...,T.$

Define the error terms $e_{tn} \equiv \ln p_{tn} - \ln\pi_t^* - \ln\alpha_n^*$ for t = 1,...,T and n = 1,...,N. If all $e_{tn} = 0$, then P^{t*} will equal P^{t**} and Q^{t*} will equal Q^{t**} for t = 1,...,T. However, if the error terms are not all equal to zero, then the statistical agency will have to decide on pragmatic grounds which option to choose.

It is straightforward to generalize the weighted least squares minimization problem (105) to the case where there are missing prices and quantities. As in section 6 we assume that there are N products and T time periods but not all products are purchased (or sold) in all time periods. For each period t, define the set of products n that are present in period t as $S(t) \equiv \{n: p_{tn} > 0\}$ for t = 1,2,...,T. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t: p_{tn} > 0\}$. Again, assume that these sets are not empty; i.e., each product is sold in at least one time period. The generalization of (105) to the case of missing products is the following weighted least squares minimization problem:

(112) $\min_{\rho,\beta} \Sigma_{t=1}^T \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho,\beta} \Sigma_{n=1}^N \Sigma_{t\in S^*(n)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2.$

Note that there are two equivalent ways of writing the least squares minimization problem. The first order necessary conditions for $\rho_1,...,\rho_T$ and $\beta_1,...,\beta_N$ to solve (112) are the following counterparts to (106) and (107):

(113) $\Sigma_{n\in S(t)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{n\in S(t)} s_{tn}\ln p_{tn}$; $t = 1,...,T;$
(114) $\Sigma_{t\in S^*(n)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{t\in S^*(n)} s_{tn}\ln p_{tn}$; $n = 1,...,N.$

As usual, the solution to (113) and (114) is not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (123) and (124), then so do $[\rho_1^*+\lambda,...,\rho_T^*+\lambda]$ and $[\beta_1^*-\lambda,...,\beta_N^*-\lambda]$ for all λ. Thus we can set $\rho_1^* = 0$ in equations (114) and drop the first equation in (113) and use linear algebra to find a unique solution for the resulting equations.

Define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* by definitions (108). Substitute these definitions into equations (113) and (114). After some rearrangement, equations (113) and (114) become the following ones:

(115) $\pi_t^* = \exp[\Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^*)]$; $t = 1,...,T;$
(116) $\alpha_n^* = \exp[\Sigma_{n\in S^*(n)} s_{tn}\ln(p_{tn}/\pi_t^*)/\Sigma_{n\in S^*(n)} s_{tn}]$; $n = 1,...,N.$

Once the estimates for the π_t and α_n have been computed, we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T; see (110) and (111) above.[89]

---

[89] The counterparts to definitions (110) are now: $P^{t*} \equiv \pi_t^* = \Pi_{n\in S(t)} \exp[s_{tn}\ln(p_{tn}/\alpha_n^*)]$, a share weighted geometric mean of the quality adjusted prices present in period t, and $Q^{t*} \equiv \Sigma_{n\in S(t)}p_{tn}q_{tn}/P^{t*}$ for t = 1,...,T. The counterparts to equations (111) are now: $Q^{t**} \equiv \Sigma_{n\in S(t)}\ \alpha_n^*q_{tn}$ and $P^{t**} \equiv \Sigma_{n\in S(t)}p_{tn}q_{tn}/Q^{t**} =$

The new price levels π_t^* defined by (115) are functions of the T price vectors, $p^1,...,p^T$ and the t quantity vectors $q^1,...,q^T$. If there are missing products, these vectors will reflect this and will not be of full dimension N. The new price level functions have the same axiomatic properties (i)-(iv) which were noted earlier in this section.[90] The present price level functions do take the economic importance of the products into account and thus are a clear improvement over their unweighted counterparts which were discussed in section 6. If the estimated errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ that implicitly appear in the weighted least squares minimization problem (112) turn out to be small, then the underlying exact model, $p_{tn} = \pi_t\alpha_n$ for $t = 1,...,T$, $n\in S(t)$, provides a good approximation to reality and thus this weighted time product dummy hedonic regression model can be used with some confidence. If the fit of the model is not good, then it may be necessary to look at other models such as those to be considered in subsequent sections.

However, the present price level functions have the same unsatisfactory property that their unweighted counterparts had in section 6: a product that is available only in one period out of the T periods has no influence on the aggregate price levels π_t^*.[91] This means that the price of a new product that appears in period T has no influence on the price levels and the price indexes generated by this model, where the price indexes are ratios of the price levels. The hedonic regression models in the next section that make use of information on the characteristics of the products do not have this unsatisfactory property of the time dummy hedonic regression models studied in sections 6-9.

10. The Time Dummy Hedonic Regression Model with Characteristics Information

In this section, it is again assumed that there are N products that are available over a window of T periods. As in the previous sections, we again assume that the quantity aggregator function for the N products is the linear function, $Q(q) \equiv \alpha\cdot q = \Sigma_{n=1}^N \alpha_n q_n$ where q_n is the quantity of product n purchased or sold in the period under consideration and α_n is the quality adjustment factor for product n. What is new is the assumption that the quality adjustment factors are functions of a vector of K *characteristics* of the

---

$\Sigma_{n\in S(t)}p_{tn}q_{tn}/\Sigma_{n\in S(t)}\ \alpha_n^* q_{tn} = \Sigma_{n\in S(t)}p_{tn}q_{tn}/\Sigma_{n\in S(t)}\ \alpha_n^*\ (p_{tn})^{-1}p_{tn}q_{tn} = [\Sigma_{n\in S(t)}\ s_{tn}(p_{tn}/\alpha_n^*)^{-1}]^{-1}$, a share weighted harmonic mean of the quality adjusted prices present in period t. Thus using Schlömilch's inequality (see Hardy, Littlewood and Polyá (1934; 26)), we see that $P^{t**} \le P^{t*}$. Similarly, $P^{1**} \le P^{1*}$. This algebra is due to de Haan (2004b) (2010) and de Haan and Krsinich (2018; 763). If the variance of prices increases over time, it is likely that P^{t**}/P^{1**} will be less that P^{t*}/P^{1*} and vice versa if the variance of prices decreases; see de Haan and Krsinich (2018; 771) and Diewert (2018;10) on this last point. Note that the work of de Haan and Krsinich provides us with a concrete formula for the difference between P^{t*} and P^{t**}. The model used by de Haan and Krsinich is a more general hedonic regression model which includes the time dummy model used in the present section as a special case. Thus their algebra noted in this footnote can be applied to all of the subsequent hedonic regression models in the following two sections that use time dummies, share weights and log prices.

[90] However, we would like the P^{t*} to satisfy a strong identity test as noted above; i.e., we would like P^{t*} to equal P^{r*} if the prices in periods t and r are identical. The $P^{t*} \equiv \pi_t^*$ where the π_t^* are defined by (115) do not satisfy this strong identity test for price levels. However, the Q^{t**} defined as $\Sigma_{n\in S(t)}\ \alpha_n^* q_{tn}$ do satisfy the strong identity test for quantities and this suggests that the P^{t**}, Q^{t**} decomposition of period t sales may be a better choice than the P^{t*}, Q^{t*} decomposition.

[91] See Diewert (2004) for a proof or modify the proof in section 6 above.

products. Thus it is assumed that product n has the vector of characteristics $z^n \equiv [z_{n1},z_{n2},...,z_{nK}]$ for n = 1,...,N. We assume that this information on the characteristics of each product has been collected.[92] The new assumption in this section is that the quality adjustment factors α_n are functions of the vector of characteristics z^n for each product and the same function, f(z) can be used for each quality adjustment factor; i.e., we have the following assumptions:

$$(117)\ \alpha_n \equiv f(z^n) = f(z_{n1},z_{n2},...,z_{nK})\ ; \qquad n = 1,...,N.$$

Thus each product has its own unique mix of characteristics but the same function f can be used to determine the relative utility to purchasers of the products. Define the period t quantity vector as $q^t = [q_{t1},...,q_{tN}]$ for t = 1,...,T. If product n is missing in period t, then define $q_{tn} \equiv 0$. Using the above assumptions, the aggregate quantity level Q^t for period t is defined as:

$$(118)\ Q^t \equiv \alpha\cdot q^t = \Sigma_{n=1}^{N} f(z^n)q_{tn}\ ; \qquad t = 1,...,T.$$

Using our assumption of (exact) utility maximizing behavior with the linear utility function defined by (118), equations (10) become the following equations:

$$(119)\ p_{tn} = \pi_t f(z^n)\ ; \qquad t = 1,...,T;\ n\in S(t).$$

The assumption of approximate utility maximizing behavior is more realistic so error terms need to be appended to equations (119). We also need to choose a functional form for the *quality adjustment* (or *hedonic*) *function* f(z). Consider the following functional form for the hedonic valuation function:

$$(120)\ f(z) = f(z_1,...,z_K) \equiv e^{\gamma_0}\prod_{k=1}^{K} z_k^{\gamma_k}\ .$$

Define the logarithms of the *quality adjustment factors* α_n as follows:

$$(121)\ \beta_n \equiv \ln\alpha_n = \ln f(z^n) = \gamma_0 + \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}\ ; \qquad n = 1,...,N$$

where we have used assumptions (117) and (120). Now take logarithms of both sides of equations (119) and add error terms e_{tn} to the resulting equations. Using equations (121), we obtain the following system of estimating equations:[93]

[92] Basically, we want to collect information on the most important price determining characteristics of each product; see Triplett (2004) and Aizcorbe (2014) for many examples of this type of hedonic regression and references to the applied literature on this topic.

[93] If both sides of equation tn in equations (122) are differentiated with respect to $\ln z_{nk}$, we find that $\partial \ln p_{tn}/\partial \ln p_{tn} = \gamma_k$ for $n\in S(t)$. Thus γ_k is the percentage change in the price of a product with respect to a one percent increase in the amount of characteristic k in a product. In general, this (constant) elasticity will be positive; i.e., a small increase in the amount of characteristic k that is present in a generic product will increase the price of the product.

(122) $\ln p_{tn} = \rho_t + \gamma_0 + \Sigma_{k=1}^{K} \gamma_k \ln z_{nk} + e_{tn}$; $t = 1,...,T;\ n \in S(t)$

where as usual, we have defined $\rho_t \equiv \ln\pi_t$ for t = 1,...,T.

Equations (122) are the equations which characterize the classic log linear time dummy hedonic regression model.[94] Note that our derivation of this model rests on the assumption of approximate utility maximizing behavior on the part of purchasers of the N products. Note also that our underlying economic model assumes that the N products are perfect substitutes once they have been quality adjusted where the quality adjustment factors are defined by (117).[95]

Estimates for $\rho \equiv [\rho_1,...,\rho_T]$ and $\gamma \equiv [\gamma_0,\gamma_1,...,\gamma_K]$ can be obtained by minimizing the sum of the squared errors e_{tn} which appear in equations (122). This leads to the following least squares minimization problem:

(123) $\min_{\rho,\gamma} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]^2$.

A solution ρ, γ to the minimization problem (123) will satisfy the following first order conditions:

(124) $\Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0$; $t = 1,...,T;$
(125) $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0$;
(126) $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]\ln z_{nk} = 0$; $k = 1,...,K.$

Equations (124)-(126) are T+1+K equations in the T+1+K unknown parameters in the vectors ρ and γ. However, solutions to these equations are not unique; if ρ_t for t = 1,...,T and γ_k for k = 0,1,...,K is a solution to (124)-(126), then $\rho_t + \lambda$ for t = 1,...,T, $\gamma_0 - \lambda$ and γ_k for k = 1,...,K is also a solution for any number λ. Thus a normalization on these parameters is required for a unique solution to (124)-(126).[96] Choose the normalization $\rho_1^* = 0$ which is equivalent to $\pi_1^* = 1$. Thus set $\rho_1^* = 0$ in equations (124)-(126), drop the first equation in equations (124) and solve the remaining T+K equations for $\rho_2^*,..., \rho_T^*$ and $\gamma_0^*, \gamma_1^*,..., \gamma_K^*$.[97] Once these parameters have been determined, the estimated $\beta_n^* \equiv \ln\alpha_n^*$ can be defined as follows using definitions (121):

(127) $\beta_n^* \equiv \ln\alpha_n^* = \ln f(z^n) = \gamma_0^* + \Sigma_{k=1}^{K} \gamma_k^* \ln z_{nk}$; $n = 1,...,N.$

[94] This model was first introduced by Court (1939) as his hedonic suggestion number 2. It was popularized by Griliches (1971; 7). See Triplett (2004) and Aizcorbe (2014) for hundreds of references to the literature on the use of this model.

[95] Thus smaller in magnitude errors e_{tn} in the hedonic regression imply that the underlying economic model provides a closer approximation to actual behavior; i.e., a higher R^2 for the linear regression model defined by (122) means that the underlying economic model provides a closer approximation to behavior.

[96] As usual, we also need the modified equations (124)-(126) to satisfy a full rank condition so that the matrix of coefficients associated with these equations can be inverted. Thus in particular, K, the number of characteristics, cannot be to big relative to N, the number of products.

[97] Alternatively, set $\rho_1 = 0$ in equations (122) and run a simple linear regression to obtain estimates for the remaining parameters.

Using equations (124) evaluated at ρ^* and γ^* and definitions (137), we see that $\ln\pi_t^* \equiv \rho_t^*$ is equal to the following expression:

(128) $\ln\pi_t^* = [1/N(t)]\ \Sigma_{n\in S(t)} \ln(p_{tn}/\alpha_n^*)$; $\qquad$ t = 1,...,T

where $\alpha_n^* \equiv \exp[\beta_n^*]$ for n = 1,...,N and where N(t) is equal to the number of products that are available in period t. Thus the estimated period t price level, π_t^*, is an *equally weighted geometric average of the quality adjusted prices* p_{tn}/α_n^* for the products that are present in period t.[98] Once the π_t^* have been calculated, the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \leq t, \tau \leq T$. If quantity data are available, then we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T; see (110) and (111) above.

It is useful to compare the present time dummy hedonic regression that uses characteristics information with the time dummy product regression in section 6 where the only characteristic of each product was the product itself. It seems that this earlier model is more general than the present model; i.e., define β_n^* by definitions (127) for n = 1,...,N). Substitute these β_n^* into the objective function for the minimization problem defined by (42) in section 6. Thus these β_n^* are feasible β_n that could be inserted into (42) but they may not be optimal; i.e., in general, we can expect the time dummy product least squares minimization problem defined by (42) to deliver a *lower* sum of squared residuals than the solution to (123) delivers. Thus we might ask at this point why consider the least squares problem (123) when in general, the least squares problem (42) will deliver a better outcome in terms of fitting the data? The problem with (42) is that there may be *no solution* to the least squares minimization problem if product turnover is rapid; i.e., if there are very few matched models in the window of observations, then the regression associated with (42) may not have enough degrees of freedom to provide a solution to the first order condition equations that are associated with this model. An extreme case where there is no solution to (42) is the case where every product is a new one which appears in only one period.[99] Thus the use of hedonic regressions with characteristics information is particularly useful in situations where there is rapid product turnover and there are very few matched models.

The price levels π_t^* defined by (128) are not entirely satisfactory for the following reason: suppose periods τ and t have exactly the same set of products that are available for those two periods. Then the price index between those two periods is equal to the following expression:

[98] An equivalent result was derived in Triplett and McDonald (1977; 150).

[99] Housing is an example of such a unique product. Every dwelling unit is uniquely determined by its location and over time, the structure associated with the housing unit depreciates in value with age (or it may appreciate in value due to renovations and improvements). Thus hedonic regressions with housing characteristics information must be used in order to obtain useful price indexes for housing. For applications of hedonic regressions to property prices, see Eurostat (2013), Diewert, Haan and Hendricks (2015), Hill (2013), Diewert and Shimizu (2016) (2017), Diewert, Huang and Burnett-Issacs (2017) and Silver (2018).

$$(129)\ \pi_t^*/\pi_\tau^* = \prod_{n\in S(t)} (p_{tn}/p_{\tau n})^{1/(N(t)}.$$

Thus the price index between the two periods is equal to a simple (equally weighted) geometric average of the price ratios $p_{tn}/p_{\tau n}$ for the products that are present in both periods; i.e., the economic importance of the products is not taken into account.[100]

In previous sections, we noted that weighting prices by their economic importance was generally recommended (but not necessary if the fit of the corresponding hedonic regression was good). The same conclusion applies in the present context. Thus if quantity information is available (in addition to price and product characteristic information), then it is preferable to generate ρ and γ estimates by solving the following *weighted least squares minimization problem*:[101]

$$(130)\ \min_{\rho,\gamma} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]^2$$

where the expenditure or sales shares s_{tn} are defined as $s_{tn} \equiv p_{tn}q_{tn}/\Sigma_{i\in S(t)} p_{ti}q_{ti}$ for $t = 1,...,T$ and $n\in S(t)$. A solution ρ, γ to the minimization problem (130) will satisfy the following first order conditions:

$$(131)\ \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0\ ; \qquad t = 1,...,T;$$
$$(132)\ \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0\ ;$$
$$(133)\ \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]\ln z_{nk} = 0\ ; \qquad k = 1,...,K.$$

Equations (131)-(132) are T+1+K equations in the T+1+K unknown parameters in the vectors ρ and γ. However, solutions to these equations are not unique; if ρ_t for $t = 1,...,T$ and γ_k for $k = 0,1,...,K$ is a solution to (124)-(126), then $\rho_t + \lambda$ for $t = 1,...,T$, $\gamma_0 - \lambda$ and γ_k for $k = 1,...,K$ is also a solution for any number λ. Thus a normalization on these parameters is required for a unique solution to (131)-(132).[102] Choose the normalization $\rho_1^* = 0$ which is equivalent to $\pi_1^* = 1$. Thus set $\rho_1^* = 0$ in equations (131)-(133), drop the first equation in equations (131) and solve the remaining T+K equations for $\rho_2^*,..., \rho_T^*$ and $\gamma_0^*, \gamma_1^*,..., \gamma_K^*$. Once these parameters have been determined, the estimated $\beta_n^* \equiv \ln\alpha_n^*$ can be defined as $\beta_n^* \equiv \gamma_0^* + \Sigma_{k=1}^{K} \gamma_k^* \ln z_{nk}$ for $n = 1,...,N$. Once the β_n^* have been defined, the corresponding quality adjustment factors are defined as $\alpha_n^* \equiv \exp[\beta_n^*] > 0$ for $n = 1,...,N$.

---

[100] As in section 6, we note that if the estimated squared residuals for this model are small, then the estimated π_t^* defined by (128) will be satisfactory since in this case, $p^t \approx \pi_t^*\alpha^*$ so that prices vary (approximately) proportionally over time and thus $\prod_{n=1}^{N} (p_{tn}/\alpha_n^*)^{1/N} \approx \pi_t^*$ for $t = 1,...,T$. Any missing price for period t and product n is defined as $p_{tn} \equiv \pi_t^*\alpha_n^*$ in the products $\prod_{n=1}^{N} (p_{tn}/\alpha_n^*)^{1/N}$. The idea of using the R^2 or the fit of a hedonic regression model to judge its adequacy can be traced back to Silver (2010; S220) (2011; 561). He implicitly suggested that hedonic regressions should only be used when the products under consideration are highly substitutable and hence when the R^2 for the relevant hedonic regression is high.

[101] Diewert (2003b) (2005b) considered this model for the bilateral case where T = 2. Silver and Heravi (2005) considered the general model.

[102] As usual, we need a full rank condition to be satisfied so that the matrix of coefficients in the system of linear equations involving ρ and γ can be inverted.

Using equations (131) evaluated at ρ^* and γ^*, we see that $\pi_t^* \equiv \exp[\rho_t^*]$ is equal to the following expression:[103]

(134) $\pi_t^* = \exp[\Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^*)]$; $\qquad$ t = 1,...,T

with $\pi_1^* \equiv 1$. Thus the period t estimated price level, π_t^*, is an expenditure share weighted geometric mean of the quality adjusted period t prices, p_{tn}/α_n^*, for the products n that are present in period t. Once the π_t^* have been calculated, the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \leq t, \tau \leq T$. If quantity data are available, then we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T; see (110) and (111) above.

The new price indexes are a clear improvement over their unweighted counterparts defined earlier by equation (129). In the present situation, using equations (134), we see that π_t^*/π_τ^* is a share weighted geometric mean of the quality adjusted period t prices, p_{tn}/α_n^*, for the products n that are present in period t with weights s_{tn} in the numerator divided by the share weighted geometric mean of the quality adjusted period τ prices, $p_{\tau n}/\alpha_n^*$, for the products n that are present in period τ with weights $s_{\tau n}$ in the denominator. Thus economic importance counts in the present model whereas it did not in the corresponding unweighted model.

Using the solution functions for the price levels π_t^* given by (134) plus the definition of the weighted least squares minimization problem given by (13), it can be shown that π_t^*/π_τ^* regarded as a function of $p^1,...,p^T$ and $q^1,...,q^T$ satisfies the following tests: (i) the weak identity test; i.e., $\pi_t^*/\pi_\tau^* = 1$ if $p^\tau = p^t$ *and* $q^\tau = q^t$ (ii) time reversal; i.e., $\pi_t^*/\pi_\tau^* = 1/[\pi_\tau^*/\pi_t^*]$;[104] (iii) homogeneity of degree 1 in period t prices; i.e., if we regard π_t^*/π_τ^* as a function of the vector of period t prices, p^t, by looking at (130), it can be seen that $\pi_t^*(\lambda p^t) = \lambda\pi_t^*(p^t)$ and $\pi_\tau^*(\lambda p^t) = \pi_\tau^*(p^t)$ for all $\lambda > 0$ which establishes the desired property; (iii) homogeneity of degree –1 in period τ prices; i.e., $\pi_t^*(\lambda p^\tau) = \pi_t^*(p^\tau)$ and $\pi_\tau^*(\lambda p^\tau) = \lambda\pi_\tau^*(p^\tau)$ for all $\lambda > 0$ which establishes the desired property; (iv) homogeneity of degree 0 in period t quantities; i.e., π_t^*/π_τ^* does not change if q^t is replaced by λq^t for all $\lambda > 0$; (v) homogeneity of degree 0 in period τ quantities; i.e., ., π_t^*/π_τ^* does not change if q^τ is replaced by λq^τ for all $\lambda > 0$; (vi) invariance to changes in the units of measurement for both products and characteristics and (vii) invariance to the ordering of the products.

Recall that the weighted time product dummy price levels defined in the previous section had the undesirable property that a product that is available only in one period out of the T periods had no influence on the aggregate price levels π_t^*. This meant that the price of a new product that appears in period T had no influence on the resulting price levels. The weighted time dummy hedonic regressions defined in this section no longer have this undesirable property.

[103] These equations are equivalent to equations (8) in de Haan and Krsinich (2018; 760).
[104] See Diewert (2005b; 773) for similar proofs.

Recall that the estimated quality adjustment factors for the N products in the model are the α_n^* for n = 1,...,N. The logarithms of these estimated quality adjustment factors are the β_n^* defined by definitions (127). Once the $\rho^* \equiv [\rho_1^*,\rho_2^*,...,\rho_T^*]$ and $\gamma^* \equiv [\gamma_0^*,\gamma_1^*,...,\gamma_K^*]$ solution to (130) has been determined (with $\rho_1^* = 1$), we can use equations (122) to determine the estimated residuals e_{tn}^* for the model defined by (130). Thus we have the following equations:

$$(135)\ e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \gamma_0^* - \Sigma_{k=1}^K \gamma_k^* \ln z_{nk}\ ; \qquad t = 1,...,T;\ n\in S(t)$$
$$= \ln p_{tn} - \rho_t^* - \beta_n^* \qquad \text{using definitions (127)}$$
$$= \ln(p_{tn}/\pi_t^*) - \beta_n^* \qquad \text{since } \rho_t^* \equiv \ln\pi_t^*.$$

Using definitions (135), it can be seen that the β_n^* satisfy the following equations:

$$(136)\ \beta_n^* = \ln(p_{tn}/\pi_t^*) - e_{tn}^*\ ; \qquad n = 1,...,N;\ t\in S^*(n).$$

For each n, consider the following share weighted average of the β_n^* that appear in equations (136):

$$(137)\ \Sigma_{t\in S^*(n)}\, s_{tn}\beta_n^* = \Sigma_{t\in S^*(n)}\, s_{tn}[\ln(p_{tn}/\pi_t^*) - e_{tn}^*]\ ; \qquad n = 1,...,N$$
$$\approx \Sigma_{t\in S^*(n)}\, s_{tn}\ln(p_{tn}/\pi_t^*)$$

since the minimization problem defined by (130) will make the squared errors $(e_{tn}^*)^2$ small within the constraints of the hedonic model. Thus we have the following approximation for the β_n^*:[105]

$$(139)\ \beta_n^* \approx [\Sigma_{t\in S^*(n)}\, s_{tn}\ln(p_{tn}/\pi_t^*)]/\Sigma_{t\in S^*(n)}\, s_{tn}\ ; \qquad n = 1,...,N.$$

Thus the logarithm of the product n quality adjustment factor, β_n^*, is approximately equal to a share weighted average of the logarithms of the inflation adjusted prices p_{tn}/π_t^* for product n over the periods t when this product was sold (or purchased) on the marketplace. Note that the averages on the right hand sides of the approximate equalities (139) are taken over the entire sample period.

Equations (136) and the equations $\beta_n^* \equiv \ln\alpha_n^*$ for n = 1,...,N can be used to establish the following equations:

$$(140)\ p_{tn} = \alpha_n^*\pi_t^* \exp[e_{tn}^*]\ ; \qquad t = 1,...,T.$$

Suppose that the underlying hedonic model holds exactly so that each error term e_{tn}^* is equal to 0. Suppose further that all of the products are *perfect substitutes* without quality adjustment so that all α_n equal α_1 for n = 1,...,N. Then the α_n^* will equal α_1^* for n = 1,...,N as well. Since the $e_{tn}^* = 0$, $\exp[e_{tn}^*] = 1$ for t = 1,...,T; $n\in S(t)$. Substitute these

[105] These equations provide approximate counterparts to equations (114) which were exact for the weighted time product dummy model discussed in section 9 above.

relationships into equations (140). Now multiple both sides of equation tn in equations (140) by q_{tn} for t = 1,...,T; $n \in S(t)$. We obtain the following system of equations after a certain amount of summation within each period:

$$(141)\ \Sigma_{n\in S(t)}\, p_{tn}q_{tn} = \alpha_1^*\pi_t^*\, \Sigma_{n\in S(t)}\, q_{tn}\ ; \qquad t = 1,...,T.$$

Now take ratios of equations (141) for t = 1 and a general t. After a bit of rearrangement, we obtain the following expression for the price index between periods 1 and t:

$$(142)\ \pi_t^*/\pi_1^* = \{\Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, q_{tn}\}/\Sigma_{n\in S(1)}\, p_{1n}q_{1n}/\Sigma_{n\in S(1)}\, q_{1n}\}; \qquad t = 1,...,T.$$

The right hand side of (142) for period t can be recognized as the *unit value price index* between periods 1 and t.

The above algebra helps to resolve an index number discontinuity problem recognized by de Haan and Krsinich (2018; 760). These authors noted that the weighted geometric mean representation for $\pi_t^* = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]$ (recall equations (134)) did not seem to collapse down to a unit value index if all of the estimated α_n^* were equal which is disconcerting because if the products are perfect substitutes (without quality adjustment), then the appropriate index should collapse down to a unit value index (because each additional unit of any product gives the purchaser the same utility). However, if the products are perfect substitutes and markets are functioning properly, the price of every product in the group under consideration should be the same in each period. Under these conditions, the estimated α_n^* will all be equal and equations (140) will become $p_{tn} = \alpha_1^*\pi_t^*$ and equations (142) will hold. Thus under these conditions, there is no discontinuity problem.

Once the estimated coefficients $\pi^* \equiv [\pi_1^*,...,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,...,\alpha_N^*]$ have been determined, these estimates can be used to determine *imputed prices* for the missing observations; i.e., if product n in period t is missing, define $p_{tn} \equiv \pi_t^*\alpha_n^*$. The corresponding missing quantities and shares are defined as $q_{tn} \equiv 0$ and $s_{tn} \equiv 0$. Using these imputed prices and quantities, we can form complete price, quantity and share vectors for all N products for each period t. Denote these vectors as p^t, q^t and s^t for t = 1,...,T. Using the fact that the share for a missing product is equal to zero, we can rewrite equations (134) as follows:

$$(143)\ \pi_t^* = \Pi_{n=1}^N (p_{tn}/\alpha_n^*)^{s_{tn}}\ ; \qquad t = 1,...,T.$$

Define the sequence of *hedonic price indexes*, P_H^t, as $P_H^t \equiv \pi_t^*/\pi_1^*$ for t = 1,...,T.[106] Using equations (143) and $\beta_n^* \equiv \ln\alpha_n^*$ for n = 1,...,N, we have the following expressions for the logarithms of the hedonic price indexes:

[106] Since $\pi^{1*} = 1 = \exp[\Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^*)] = \exp[\Sigma_{n=1}^N\, s_{1n}\ln(p_{1n}/\alpha_n^*)]$, $P_H^t \equiv \pi^{t*}/\pi^{1*} = \pi^{t*}$ for t = 1,...,T. However, when we compare P_H^t to the corresponding fixed base Törnqvist index P_T^t, it proves to be more convenient to define P_H^t as π^{t*}/π^{1*} for t = 1,...,T. Note that the P_H^t do not depend on the missing prices. However, the P_T^t do depend on the missing prices.

(144) $\ln P_H^t = \Sigma_{n=1}^N s_{tn}(\ln p_{tn} - \beta_n^*) - \Sigma_{n=1}^N s_{1n}(\ln p_{1n} - \beta_n^*)$; $t = 1,...,T.$

It is now possible to compare the sequence of price indexes to the corresponding Törnqvist fixed base indexes that make use of the imputed prices generated by the present model for the missing products. The logarithm of the *fixed base Törnqvist price index* between periods 1 and t, P_T^t, is defined as follows:

(145) $\ln P_T^t \equiv \Sigma_{n=1}^N ½ (s_{tn} + s_{1n})(\ln p_{tn} - \ln p_{1n})$ $t = 1,...,T$
$= \Sigma_{n=1}^N ½ (s_{tn} + s_{1n})[\ln p_{tn} - \beta_n^* - (\ln p_{1n} - \beta_n^*)$.

Taking the difference between (144) and (145), we can derive the following expressions for t = 1,2,...,T:[107]

(146) $\ln P_H^t - \ln P_T^t = \Sigma_{n=1}^N ½(s_{tn} - s_{1n})(\ln p_{tn} - \beta_n^*) + \Sigma_{n=1}^N ½(s_{tn} - s_{1n})(\ln p_{1n} - \beta_n^*)$.

Since $\Sigma_{n=1}^N (s_{tn} - s_{t1}) = 0$ for each t, the two sets of terms on the right hand side of equation t in (146) can be interpreted as normalizations of the covariances between $s^t - s^1$ and $\ln p^t - \beta^*$ for the first set of terms and between $s^t - s^1$ and $\ln p^1 - \beta^*$ for the second set of terms. If the products are highly substitutable with each other, then a low p_{tn} will usually imply that $\ln p_{tn}$ is less than the average log price β_n^* and it is also likely that s_{tn} is greater than s_{1n} so that $(s_{tn} - s_{1n})(\ln p_{tn} - \beta_n^*)$ is likely to be negative. Hence the covariance between $s^t - s^1$ and $\ln p^t - \beta^*$ will tend to be negative. On the other hand, if p_{1n} is unusually low, then $\ln p_{tn}$ will be less than the average log price β_n^* and it is likely that s_{1n} is greater than s_{tn} so that $(s_{tn} - s_{1n})(\ln p_{1n} - \beta_n^*)$ is likely to be positive. Hence the covariance between $s^t - s^1$ and $\ln p^1 - \beta^*$ will tend to be positive. Thus the first set of terms on the right hand side of (146) will tend to be negative while the second set will tend to be positive. If there are no divergent trends in log prices and sales shares, then it is likely that these two terms will largely offset each other and under these conditions, P_H^t is likely to approximate P_T^t reasonably well. However, with divergent trends and highly substitutable products, it is likely that the first set of negative terms will be larger in magnitude than the second set of terms and thus P_{WTPD}^t is likely to be below P_T^t under these conditions. On the other hand, if there are missing products in period 1, then the second set of covariance terms can become very large and positive and outweigh the first set of generally negative terms.[108] The bottom line is that P_H^t and P_T^t can diverge substantially. In such a case, it may be preferable to use the hedonic regression to simply fill in the missing prices and use a superlative index to generate price indexes rather than use the price levels π_t^* generated by the hedonic time dummy regression as the price indexes.[109]

[107] The analysis here follows that of Diewert (2018; 52-53).
[108] See Diewert (2018; 39) for just such an example.
[109] However, if the fit in the hedonic regression is good, then prices are close to being proportional over time and the price levels generated by the hedonic regression will generate satisfactory results.

The hedonic valuation function f(z) defined by (120) has a useful property: one can impose constant returns to scale in the characteristics(the property $f(\lambda z) = \lambda f(z)$ for all $\lambda > 0$) if the γ_k satisfy the restriction $\Sigma_{k=1}^{K} \gamma_k = 1$. However, if we want to apply equations (121) as estimating equations for the unknown parameters in f(z), we need *positive amounts of all characteristics in all models* so that $\ln z_{nk}$ is well defined; i.e., we need $z_{nk} > 0$ for all n = 1,...,N and k = 1,...,K. The first alternative hedonic regression model to be considered in the following section relaxes this positivity restriction.

11. Alternative Hedonic Regression Models with Characteristics Information

As noted in the previous section, the hedonic valuation function f(z) defined by (120) requires that positive amounts of all characteristics be present in all N models. It would be useful to have a hedonic regression model that could in principle deal with the introduction of new characteristics over the sample period. This can be done if we replace the f(z) defined by (120) by the following functional form for f(z):

(147) $\ln f(z_1,z_2,...,z_K) \equiv \gamma_0 + \Sigma_{k=1}^{K} \gamma_k z_k$.

Using this new hedonic valuation function and making the same assumptions (117)-(119) as were made in the previous section along with the new assumption (147), we obtain the following system of estimating equations which are counterparts to equations (122):

(148) $\ln p_{tn} = \rho_t + \gamma_0 + \Sigma_{k=1}^{K} \gamma_k z_{nk} + e_{tn}$; $\quad t = 1,...,T; n \in S(t)$

where as usual, $\rho_t \equiv \ln \pi_t$ for t = 1,...,T. We again find estimators for the unknown parameters in equations (148) by minimizing the following sum of weighted squared residuals e_{tn}:[110]

(149) $\min_{\rho, \gamma} \Sigma_{t=1}^{T} \Sigma_{n \in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k z_{nk}]^2$.

A solution ρ, γ to the minimization problem (149) will satisfy the first order conditions (131)-(133) in the previous section, except that z_{nk} replaces $\ln z_{nk}$ for all n and k. The rest of the analysis of the hedonic regression model defined by (149) follows along the same lines as the share weighted model in the previous section. Equations (127) and (135) are replaced by the following equations:

(150) $\beta_n^* \equiv \ln \alpha_n^* = \ln f(z^n) = \gamma_0^* + \Sigma_{k=1}^{K} \gamma_k^* z_{nk}$; $\quad n = 1,...,N;$
(151) $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \gamma_0^* - \Sigma_{k=1}^{K} \gamma_k^* z_{nk}$; $\quad t = 1,...,T; n \in S(t).$

The remaining equations (134)-(143) in section 10 apply to the model defined by (149). Once the π_t^* have been calculated, the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \leq t, \tau \leq T$.

---

[110] This is precisely the model studied by de Haan and Krsinich (2018). The results we derive below are identical to their results.

As usual, we have two alternative methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T. The counterparts to definitions (110) are now: $P^{t*} \equiv \pi_t^* = \Pi_{n\in S(t)} \exp[s_{tn}\ln(p_{tn}/\alpha_n^*)]$, a share weighted geometric mean of the quality adjusted prices present in period t where the α_n^* are defined by (150), and $Q^{t*} \equiv \Sigma_{n\in S(t)}p_{tn}q_{tn}/P^{t*}$ for t = 1,...,T. The counterparts to equations (111) are now: $Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^*q_{tn}$ and $P^{t**} \equiv \Sigma_{n\in S(t)}p_{tn}q_{tn}/Q^{t**} = \Sigma_{n\in S(t)}p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*q_{tn} = \Sigma_{n\in S(t)}p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*(p_{tn})^{-1}p_{tn}q_{tn} = [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_n^*)^{-1}]^{-1}$, a share weighted harmonic mean of the quality adjusted prices present in period t. As noted earlier, these results are due to de Haan and Krsinich (2018).

The hedonic regressions model defined by (149) (and its equally weighted version) were implemented by de Haan and Krsinich (2018) using monthly New Zealand data over 3 years (so that T = 36) for the following 7 classes of electronic products: desktop computers, laptop computers, portable media players, DVD players, digital cameras, camcorders and televisions. For each product class, they had data on approximately 40 characteristics. The data were aggregated across outlets and basically covered the New Zealand market. New products entered each of the 7 markets at monthly rates that ranged from 24% to 29% and old products disappeared at rates that ranged from 23% to 29%. Thus there was a tremendous amount of product churn in each of the 7 categories. Once the weighted and unweighted regressions defined by (149) were run for each category, the alternative price levels, P^{t*} and P^{t**}, were computed for each of the 7 categories and compared.[111] They found that P^{t*} was very close to P^{t**} for each category when the weighted regressions were used. This suggests that it may not matter that much which method for computing the P^t is used, since the direct hedonic regression price level estimates π^{t*} were always very close to the indirect estimates based on deflating period t values by $\Sigma_{n\in S(t)}\, \alpha_n^*q_{tn}$. This is a very encouraging result. However, it was a different story for the unweighted hedonic regressions: they were much more volatile than their weighted counterparts and the direct and indirect price levels that they generated were frequently noticeably different. Moreover the unweighted regressions generated a sequence of price levels that had substantially different trends than the corresponding trends for the weighed regressions. Our conclusion is that the results obtained by de Haan and Krsinich support the use of weighted hedonic regressions over their unweighted counterparts.

The above results were for regressions that covered the entire sample period. Statistical agencies that produce consumer price indexes need to produce monthly indexes that do not revise the data for the previous months. In order to deal with these constraints, Ivancic, Diewert and Fox (2009) suggested the use of a rolling window time dummy regression approach with a window length of 13 months (so that strongly seasonal

[111] The average unadjusted R^2 for the 7 weighted models was 0.981. The corresponding R^2 for the equally weighted models was 0.885. This suggests that the popular products were close substitutes with each other while the unpopular models were not as close substitutes. The fact that the R squares for the 7 classes of products were so high means that the underlying assumption of a linear aggregator function (after quality adjustment) is adequate to describe the data and thus it is not necessary to explore the alternative models for estimating reservation prices that will be explained in subsequent sections. Of course, the drawback to the hedonic regression models with characteristics is that it is necessary to collect information on characteristics whereas the reservation price models which follow do not require information on characteristics.

commodities could play a role in the resulting indexes). De Haan and Krsinich (2018; 773) implemented this rolling window approach for their seven product categories with a window length of 13 consecutive months for each weighted hedonic regression. The month to month change in the estimated price levels (using the P^{t**} option) in the last regression was used to update the results of the previous regression. Thus in the end, they could compare this rolling window approach to the generation of a price level series for each of the 7 categories with the corresponding one big weighed regression approach. For three of the seven categories, they found that the rolling window series ended up well below the corresponding single regression series and for one category, the rolling window series ended up well above the corresponding single regression series. This is evidence of chain drift in these four rolling window series. For these four series, it may be best to lengthen the window length for the rolling window hedonic regressions. This will usually cure the chain drift problem.[112] Statistical agencies should study this paper very carefully.

For our next hedonic model, we introduce a *discrete characteristic category*; i.e., each product n has an x characteristic where there are M separate states for this characteristic. For example, the product may come in 3 distinct package sizes: small, medium and large. In this case, M = 3. In addition, there are K price determining continuous characteristics and each product n has varying amounts of these characteristics. As usual, denote the vector of continuous characteristics for product n by $z^n \equiv [z_{n1},...,z_{nK}]$ for n = 1,...,N. If product n belongs to discrete category m, define the M dimensional vector x^n for this product as $x^n \equiv [x_{n1},...,x_{nM}] = e^m$ where e^m is a unit vector with a 1 in component m and zeros elsewhere. We assume that there is at least one product that belongs to each of the M discrete categories. We assume the existence of a hedonic valuation function, $f(z^n,x^n)$, that gives us the relative values for the N products where the logarithm of $f(z^n,x^n)$ is defined as follows:

$$(152)\ \ln f(z^n,x^n) \equiv \gamma_0 + \Sigma_{k=1}^K \gamma_k z_{nk} + \Sigma_{m=1}^M \delta_m x_{nm} ; \qquad n = 1,...,N.$$

As usual, the exact hedonic model for the prices is $p_{tn} = \pi_t f(z^n,x^n)$ for t = 1,...,T and $n \in S(t)$. Upon taking logarithms of both sides of these price equations, using $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and using definitions (152), we obtain the following hedonic regression model:

$$(153)\ \ln p_{tn} = \rho_t + \gamma_0 + \Sigma_{k=1}^K \gamma_k z_{nk} + \Sigma_{m=1}^M \delta_m x_{nm} + e_{tn} ; \qquad t = 1,...,T;\ n \in S(t).$$

---

[112] Suppose we estimate (149) where the window length is T. The solution for π_t^* is given by (134) which is $\pi_t^* = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]$ for t = 1,...,T. If the prices and shares for periods t and r coincide, then obviously, $\pi_t^* = \pi_r^*$ and Walsh's multiperiod identity test is satisfied and there can be no chain drift within the initial window of observations. Now add a new observation T+1 and drop the old observation 1. Compute new $\pi_t^{**} = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^{**})]$ for t = 2,3,...,T+1 where the new estimated parameters are π_t^{**} for t = 2,3,...,T+1 and α_n^{**} for n = 1,...,N. Now suppose the prices and shares of periods T+1 and t (where $2 \le t \le T$) are the same. Then $\pi_{T+1}^{**} = \pi_t^{**} \equiv \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^{**})] \approx \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^{*})] \equiv \pi_t^*$ where the approximation becomes closer the larger is T for a fixed N since for large T, α_n^{**} will be close to α_n^* if product churn is not too great. Recall equations (139) which also apply in the present context.

Rather than minimizing the sum of squared residuals, we find estimates for the unknown parameters in equations (153) by minimizing the following *weighted* sum of squared residuals:

(154) $\min_{\rho, \gamma, \delta} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k z_{nk} - \Sigma_{m=1}^{M} \delta_m x_{nm}]^2$

where $\rho \equiv [\rho_1,...,\rho_T]$, $\gamma \equiv [\gamma_0,\gamma_1,...,\gamma_K]$ and $\delta \equiv [\delta_1,...,\delta_M]$. A solution ρ, γ, δ to the minimization problem (151) will satisfy the following first order conditions:

(155) $\Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk} - \Sigma_{m=1}^{M} \delta_m x_{nm}] = 0$; $t = 1,...,T$;
(156) $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk} - \Sigma_{m=1}^{M} \delta_m x_{nm}] = 0$;
(157) $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk} - \Sigma_{m=1}^{M} \delta_m x_{nm}]z_{nk} = 0$; $k = 1,...,K$;
(158) $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk} - \Sigma_{m=1}^{M} \delta_m x_{nm}]x_{nm} = 0$; $m = 1,...,M$.

Equations (155)-(158) are T+1+K+M equations in the T+1+K+M unknown parameters in the vectors ρ, γ and δ. However, solutions to these equations are not unique: the variables associated with the ρ_t, γ_0 and the δ_m parameters are collinear. In order to obtain a unique solution to equations (155)-(158), it is necessary to impose two normalizations on these parameters. Choose the normalizations $\rho_1^* = 0$ (which is equivalent to $\pi_1^* = 1$) and $\delta_1^* = 0$. Thus set $\rho_1^* = 0$ and $\delta_1^* = 0$ in equations (155)-(158), drop the first equation in equations (155), drop the first equation in (158) and solve the remaining T+K+M−1 equations for $\rho_2^*,..., \rho_T^*, \gamma_0^*, \gamma_1^*,..., \gamma_K^*, \delta_2^*,..., \delta_M^*$.[113] Once these parameters have been determined, define the estimated *logarithms of the N quality adjustment factors* as:

(159) $\beta_n^* \equiv \gamma_0^* + \Sigma_{k=1}^{K} \gamma_k^* \ln z_{nk} + \Sigma_{m=1}^{M} \delta_m x_{nm}$; $n = 1,...,N$.

Once the β_n^* have been defined, the corresponding *quality adjustment factors* are defined as $\alpha_n^* \equiv \exp[\beta_n^*] > 0$ for $n = 1,...,N$. Evaluate equations (155)-(158) at the solution ρ^*, γ^*, δ^* where $\rho_1^* = 0$ and $\delta_1^* = 0$.[114] Using definitions (159), equations (155) evaluated at the above solution become the following equations:

(160) $\rho_t^* = \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \beta_n^*]$; $t = 1,...,T$.

Thus the period t estimated price level $\pi_t^* \equiv \exp[\rho_t^*]$ is a period t share weighted geometric average of the period t quality adjusted prices, p_{tn}/α_n^*, for $n\in S(t)$.

With some new definitions, it is possible to provide fairly transparent interpretations for the discrete variable parameters, the δ_m^*. Define the set of observations t,n that are in product group m as $S^{**}(m)$ for $m = 1,...,M$. For each model n, define the *partial log adjustment factor* μ_n^* as follows:

---

[113] The number of observations in the window of observations must be equal to or greater than T+K+M−1. More generally, the rank of the coefficient matrix that is associated with the T+K+M−1 remaining equations in the system of equations defined by (155)-(158) is assumed to be full so that the coefficient matrix has an inverse.
[114] All T+K+M+1 of the equations (155)-(158) will be satisfied at this solution.

(161) $\mu_n^* \equiv \gamma_0^* + \Sigma_{k=1}^K \gamma_k^* \ln z_{nk}$; n = 1,...,N.

Using these new definitions, it can be seen that equations (158), evaluated at the normalized solution to the weighted least squares minimization problem (154), can be rewritten as follows:

(162) $\delta_m^* = \Sigma_{t,n\in S^{**}(m)}\, s_{tn}[\ln p_{tn} - \rho_t^* - \mu_n^*]/\Sigma_{t,n\in S^{**}(m)}\, s_{tn}$; m = 1,...,M.

Define $\theta_n^* = \exp[\mu_n^*]$ for n = 1,...,N. Then $\exp[\delta_m^*]$ is equal to a share weighted geometric average of the partially quality adjusted prices $p_{tn}/\pi_t^*\theta_n^*$ for all t,n that belong to the set $S^{**}(m)$; i.e., for all observations over all periods on products that are in group m for the discrete characteristic. Thus the characterizations of the δ_m^* given by equations (162) are intuitively plausible.

Equations (134)-(143) in section 10 can be adapted to the model defined by (154). Once the π_t^* have been calculated, the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \leq t, \tau \leq T$. As usual, we have two alternative methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T; see (110) and (111) above for the two methods.

As was the case for the model at the end of section 10, the present model could be used to generate missing prices using $p_{tn} \equiv \pi_t^*\alpha_n^*$ to estimate the missing prices and then other index number formulae could be used to construct the period by period price and quantity levels using complete price and quantity vectors for all periods in the window of T periods.

In many cases, the continuous characteristics which describe a product or model range from very low values to very high values. In such cases, it is unlikely that a single parameter γ_k could provide an adequate approximation to the value of the extra amount of the characteristic over the entire range of feasible characteristic values. To deal with this difficulty, piecewise linear spline functions can be introduced into the hedonic model. Thus let y be the amount of a continuous characteristic that takes on a wide range of values. We again assume that there are N models or products and T time periods and we can observe the amounts $z_1,...,z_K$ and y that each product has.

In order to obtain more flexibility with respect to the y characteristic, the observed products are grouped into say 3 groups with respect to the amounts of y that they possess; low, medium and high amounts of y. In order to define this grouping, pick y^* and y^{**} such that approximately 1/3 of the observations have $y \leq y^*$, 1/3 have $y^* < y \leq y^{**}$ and 1/3 have $y^{**} < y$. Define the following *dummy variable functions*, $D_i(y)$ for i = 1,2,3, which depend on y:

(163) $D_1(y) \equiv 1$ if $y \leq y^*$ and is 0 elsewhere;
(164) $D_2(y) \equiv 1$ if $y^* < y \leq y^{**}$ and is 0 elsewhere;
(165) $D_1(y) \equiv 1$ if $y^{**} < y$ and is 0 elsewhere.

The above functions can be used to define the logarithm of the function g(y):

(166) $\ln g(y) \equiv D_1(y)\phi_1 y + D_2(y)[\phi_1 y^* + \phi_2(y - y^*)] + D_3(y)[\phi_1 y^* + \phi_2(y^{**} - y^*) + \phi_2(y - y^{**})]$.

Note that the logarithm of g(y) is a piecewise linear function of y.[115] If $\phi_1 = \phi_2 = \phi_2$, then $\ln g(y) = \phi_1 y$; i.e., under these conditions, lng(y) becomes a linear function of y.

We assume the existence of a *hedonic valuation function*, $f(z^n,y^n)$, that gives us the relative values for the N products where the logarithm of $f(z^n,y^n)$ is defined as follows:

(167) $\ln f(z^n,y^n) \equiv \gamma_0 + \Sigma_{k=1}^K \gamma_k z_{nk} + \ln g(y^n)$; n = 1,...,N.

As usual, the exact hedonic model for the prices is $p_{tn} = \pi_t f(z^n,x^n)$ for t = 1,...,T and $n \in S(t)$. Upon taking logarithms of both sides of these price equations, using $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and using definitions (166) and (167), we obtain the following hedonic regression model:

(168) $\ln p_{tn} = \rho_t + \gamma_0 + \Sigma_{k=1}^K \gamma_k z_{nk} + \ln g(y^n) + e_{tn}$; t = 1,...,T; $n \in S(t)$

where $\ln g(y^n)$ is defined by evaluating (166) at $y = y^n$. It can be seen that the unknown parameters, $\rho \equiv [\rho_1,...,\rho_T]$, $\gamma \equiv [\gamma_0,\gamma_1,...,\gamma_K]$ and $\phi \equiv [\phi_1,\phi_2, \phi_3]$ appear on the right hand sides of equations (168) in a linear fashion so the unknown parameters can be estimated using linear regression techniques.

Rather than minimizing the sum of squared residuals, we can find estimates for the unknown parameters in equations (168) by minimizing the following *weighted* sum of squared residuals:

(169) $\min_{\rho, \gamma, \phi} \Sigma_{t=1}^T \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^K \gamma_k z_{nk} - \ln g(y^n)]^2$.

We leave the further analysis of this model to the reader after noting that in order to obtain a unique solution to (169), we require a normalization on the ρ_t and γ_0 such as $\rho_1^* = 0$.

It is not necessary to restrict ourselves to hedonic regression models where the hedonic valuation function f(z) is such that lnf(z) is linear in the unknown parameters. One can choose functions f(z) such that lnf(z) is a nonlinear function of the unknown parameters and use nonlinear estimation techniques to estimate the parameters. However, when estimating nonlinear regression models that are fairly complex, it is not wise to attempt to estimate the final model right away. It is best if there are very simple models that can be nested in the final model so that one starts by estimating the simplest model and

[115] This function is known as a *linear spline function* in the literature on nonparametric approximations. The points y^* and y^{**} are called break points or knots. With a sufficient number of break points, any continuous function can be arbitrarily well approximated by a linear spline function.

gradually, more bells and whistles are added until one arrives at the final model. The final parameter values for a simpler model should be used as starting parameter values in the next stage model if possible.[116]

All of the models for quality adjustment that we have considered thus far, constant tastes have been assumed; i.e., the functional form for the aggregator function Q(q) and for the hedonic valuation functions $f(z^n,y^n,x^n)$ have remained constant over the sample period. In the following section, this assumption will be relaxed.

12. Hedonics and the Problem of Taste Change: Hedonic Imputation Indexes

A problem with hedonic regression models that are applied over many periods is that consumer tastes may change over time. In this section, we will outline three possible methods for dealing with the problem of taste change.

The first method that could be used to deal with taste change is to restrict the time dummy hedonic regression models to the case of two adjacent periods. Each pair of periods allows for a different set of tastes.[117] As each adjacent period time dummy regression model is run for say periods t−1 and t, the estimated price level ratio, say π_t^*/π_{t-1}^*, is used as an update factor for the price level of period t−1. Each bilateral regression will generate a set of quality adjustment factors which can be used to fill in missing prices. Over time, these quality adjustment factors will change. It can be seen that this model of taste change is slightly inconsistent over time but it does allow for taste change.

The second method for dealing with taste change is similar to the first method except instead of holding tastes constant for 2 consecutive periods, we hold tastes constant for T consecutive periods. When the data for a subsequent period becomes available, the data for the first period is dropped and the data for the new period is added to form a new window of T observations and a new time dummy hedonic regression is run. This method assumes that tastes change more slowly than the first method. This *rolling window time dummy hedonic regression model*[118] has a new problem which did not arise with the adjacent period model: how should the results of the new regression be linked to the results of the previous regression? Thus suppose the first window of observations generates the sequence of price levels, π_1^1, π_2^1,..., π_T^1 and these levels are labelled as official indexes for the first T periods. Suppose the time dummy hedonic regression for the second window generates the sequence of price levels π_2^2, π_3^2,..., π_{T+1}^2. How exactly should the official index for period T+1 be constructed? Ivancic, Diewert and Fox (2009) (2011) suggested using period T as the linking observation. Krsinich (2016; 383) called this the *movement splice* method for linking the two windows. Krsinich (2016; 383) also

[116] For examples of nonlinear hedonic models that make use of this nesting technique, see Diewert, Haan and Hendricks (2015), Diewert and Shimizu (2016) (2017) and Diewert, Huang and Burnett-Issacs (2017).
[117] This method is due to Court (1939) and popularized by Griliches (1971). It is called the adjacent period time dummy hedonic regression model.
[118] This rolling window time dummy hedonic model is due to Ivancic, Diewert and Fox (2009) and Shimizu, Nishimura and Watanabe (2010).

suggested that a better choice of the linking observation in the context of her multilateral model was t = 2 and she called this the *window splice* method. De Haan (2015; 26) suggested that the link period t should be chosen to be in the middle of the first window time span; i.e., choose t = T/2 if T is an even integer or t = (T+1)/2 if T is an odd integer. The Australian Bureau of Statistics (2016; 12) called this the *half splice* method for linking the results of the two windows. Ivancic, Diewert and Fox (2011; 33) and Diewert and Fox (2017; 18) argued that each choice of a linking period t running from t = 2 to t = T is an equally valid choice of a period to link the two sets of price levels. Thus they suggested the *mean splice*, defined as the geometric mean of all of the possible estimates for π_{T+1} using each of the T−1 possible link periods. The first 3 methods of linking one window to the next window are easy to explain to the public but the mean splice seems to be the least "risky" and follows standard statistical practice; i.e., if one has many estimators for the same thing that are equally plausible, then taking a average of these estimators is recommended. It can be seen that this model of taste change is again slightly inconsistent; the models are internally consistent within each window of observations but when we move from one window to another, this internal consistency is lost.

The third method for dealing with taste change is to simply estimate a separate hedonic regression for each time period. This method is called the *hedonic imputation method*. In order to explain this method and its connection to the adjacent period time dummy model, it is necessary to develop the algebra for both methods for the case of two time periods.

We first develop the algebra for the adjacent period time dummy hedonic regression model. Recall the model defined in the previous section by equations (147)-(148). Consider the special case of this model with only two periods so that T = 2. We reparameterize the weighted least squares minimization problem defined by (149) for the case T = 2 and consider the following equivalent problem:

(170) $\min_{\chi,\gamma} \Sigma_{t=1}^{2} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \chi_t - \Sigma_{k=1}^{K} \gamma_k z_{nk}]^2$

where $\chi \equiv [\chi_1,\chi_2]$ and $\gamma \equiv [\gamma_1,...,\gamma_K]$. Comparing (149) for T = 2 with (170), it can be seen that $\chi_1 = \rho_1 + \gamma_0 = \gamma_0$ (since we set $\rho_1 = 0$) and $\chi_2 = \rho_2 + \gamma_0$. Thus the two problems are completely equivalent once we impose the normalization $\rho_1 = 0$ on (149) for the case where T = 2. The first order conditions which determine a unique solution to (170)[119] are the following 2 + K equations:

(171) $\Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \chi_t - \Sigma_{k=1}^{K} \gamma_k z_{nk}] = 0$; t = 1,2;
(172) $\Sigma_{t=1}^{2} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \chi_t - \Sigma_{k=1}^{K} \gamma_k z_{nk}]z_{nk} = 0$; k = 1,...,K.

[119] As usual, the coefficient matrix for the unknown parameters in equations (171) and (172) must be of full rank (which is K + 2), in order to obtain a unique solution. This means that the number of observations must be equal to or greater than K + 2.

Denote the solution to (171) and (172) by $\chi^* \equiv [\chi_1^*,\chi_2^*]$ and $\gamma^* \equiv [\gamma_1^*,...,\gamma_K^*]$. The parameters γ_0 and ρ_2 which were used in our initial parameterization of the model can be recovered from the solution to (171) and (172) as follows:[120]

(173) $\gamma_0^* \equiv \chi_1^*$; $\rho_2^* \equiv \chi_2^* - \chi_1^*$.

It is convenient to redefine our hedonic valuation function $f(z_1,z_2,...,z_K)$: define the logarithm of this function as $\ln f(z_1,z_2,...,z_K) \equiv \Sigma_{k=1}^K \gamma_k z_k$. Note that the constant term γ_0 has been dropped from this definition. Thus we redefine the *product quality adjustment factors*, the previous α_n, and their logarithms, the previous $\beta_n \equiv \ln\alpha_n$, as follows:

(174) $\beta_n^* \equiv \Sigma_{k=1}^K \gamma_k^* z_{nk}$; $\alpha_n^* \equiv \exp[\beta_n^*]$; $\quad n = 1,...,N.$

Thus the newly parameterized *exact model* is:

(175) $p_{tn} = \pi_t\alpha_n$; $\quad t = 1,2; n\in S(t)$

where $\pi_t \equiv \exp[\chi_t]$ for t = 1,2. Once the solution to (171) and (172) has been determined, the errors e_{tn}^* in the exact equations defined by the logarithms of equations (175) are implicitly defined in the following equations:

(176) $\ln p_{tn} = \chi_t^* + \Sigma_{k=1}^K \gamma_k^* z_{nk} + e_{tn}^*$; $\quad t = 1,2; n\in S(t)$

Using equations (171) and definitions (173) and (174), we obtain the following expressions for ρ_2^* which is the logarithm of the price index π_2^*/π_1^* generated by the time dummy adjacent period hedonic regression model:[121]

$$\begin{aligned}(177)\ ;\ \rho_2^* &\equiv \chi_2^* - \chi_1^* \\ &= \Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^*) - \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^*) \\ &= \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}] - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}].\end{aligned}$$

This completes the algebra for the reparameterization of the time dummy adjacent period hedonic regression model. In what follows, we will develop the algebra for entirely separate hedonic regression models for each period. In the above model, the hedonic surfaces for the two periods, $\chi_1^* + \Sigma_{k=1}^K \gamma_k^* z_{nk}$ and $\chi_2^* + \Sigma_{k=1}^K \gamma_k^* z_{nk}$, differed only in their constant terms. In the following model, the hedonic surfaces can shift in a non-parallel fashion.

Consider the following two weighted least squares minimization problems:

---

[120] The new γ_k^* are equal to the old γ_k^* for k = 1,...,K.

[121] If the model defined by (175) held exactly, then $p_{1n} = \pi_1^*\alpha_n^*$ for $n\in S(1)$ and $p_{2n} = \pi_2^*\alpha_n^*$ for $n\in S(2)$. Thus $p_{1n}/\alpha_n^* = \pi_1^*$ for $n\in S(1)$ and $p_{2n}/\alpha_n^* = \pi_2^*$ for $n\in S(2)$. Thus each quality adjusted period t price, p_{tn}/α_n^* for $n\in S(t)$, is an estimator for π_t^* and a weighted geometric mean of these quality adjusted prices (where the weights sum to 1) will also be an estimator for π_t^*.

(178) $\min_{\chi,\gamma} \Sigma_{n\in S(1)} s_{1n}[\ln p_{1n} - \chi^1 - \Sigma_{k=1}^K \gamma_k^1 z_{nk}]^2$;
(179) $\min_{\chi,\gamma} \Sigma_{n\in S(2)} s_{2n}[\ln p_{2n} - \chi^2 - \Sigma_{k=1}^K \gamma_k^2 z_{nk}]^2$

where the unknown parameters in (178) are χ^1, $\gamma^1 \equiv [\gamma_1^1,...,\gamma_K^1]$ and the unknown parameters in (179) are χ^2, $\gamma^2 \equiv [\gamma_1^2,...,\gamma_K^2]$. In the previous model defined by (170), there is only one vector of γ parameters to model prices in both periods while the new models defined by (178) and (179) have separate γ parameter vectors.

The first order conditions for (178) are equations (180) and (181) below while the first order conditions for (179) are equations (182) and (183) below:

(180) $\Sigma_{n\in S(1)} s_{1n}[\ln p_{1n} - \chi^1 - \Sigma_{k=1}^K \gamma_k^1 z_{nk}] = 0$;
(181) $\Sigma_{n\in S(1)} s_{1n}[\ln p_{1n} - \chi^1 - \Sigma_{k=1}^K \gamma_k^1 z_{nk}]z_{nk} = 0$; $k = 1,...,K$;
(182) $\Sigma_{n\in S(2)} s_{2n}[\ln p_{2n} - \chi^2 - \Sigma_{k=1}^K \gamma_k^2 z_{nk}] = 0$;
(183) $\Sigma_{n\in S(2)} s_{2n}[\ln p_{2n} - \chi^2 - \Sigma_{k=1}^K \gamma_k^2 z_{nk}]z_{nk} = 0$; $k = 1,...,K$.

Let χ^{1*}, $\gamma_1^{1*},..., \gamma_K^{1*}$ solve (180) and (181) and let χ^{2*}, $\gamma_1^{2*},..., \gamma_K^{2*}$ solve (182) and (183). There are now *two sets of quality adjustment factors*, $\alpha_1^{1*},..., \alpha_N^{1*}$ for period 1 and $\alpha_1^{2*},..., \alpha_N^{2*}$ for period 2. The logarithms of these parameters are defined as follows:

(184) $\ln\alpha_n^{1*} \equiv \Sigma_{k=1}^K \gamma_k^{1*} z_{nk}$; $\ln\alpha_n^{2*} \equiv \Sigma_{k=1}^K \gamma_k^{2*} z_{nk}$; $n = 1,...,N$.

Using (180), (182) and definitions (184), we obtain the following expressions for χ^{1*} and χ^{2*} as *quality adjusted log prices* for periods 1 and 2:

(185) $\chi^{1*} = \Sigma_{n\in S(1)} s_{1n}\ln(p_{1n}/\alpha_n^{1*}) = \Sigma_{n\in S(1)} s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K \gamma_k^{1*} z_{nk}]$;
(186) $\chi^{2*} = \Sigma_{n\in S(2)} s_{2n}\ln(p_{2n}/\alpha_n^{2*}) = \Sigma_{n\in S(2)} s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K \gamma_k^{2*} z_{nk}]$.

The average measure of log price change going from period 1 to 2 using the adjacent period time dummy hedonic model was $\rho^* = \chi_2^* - \chi_1^*$; see (177) above. Note that the same quality adjustment factors, the α_n^*, were used to quality adjust prices in both periods. At first glance, we might think that an analogous measure of average constant quality log price in our new model could be defined as $\chi^{2*} - \chi^{1*}$. However, looking at (185) and (186), we see that the quality adjustment factors are not held constant in constructing this measure. The underlying exact models are now $p_{1n} = \pi_1^*\alpha_n^{1*}$ for $n\in S(1)$ and $p_{2n} = \pi_2^*\alpha_n^{2*}$ for $n\in S(2)$ where $\pi_1^* \equiv \exp[\chi^{1*}]$ and $\pi_2^* \equiv \exp[\chi^{2*}]$. Thus the period 1 quality adjusted prices, p_{1n}/α_n^{1*}, are not comparable to their period 2 counterparts, p_{2n}/α_n^{2*}, unless $\alpha_n^{1*} = \alpha_n^{2*}$. Thus π_2^*/π_1^* is not a useful price index that compares like with like.

At this point, the analysis could go in at least 3 different directions:

- Use the two hedonic regressions to fill in the missing prices; i.e., if $n\in S(1)$ but $n\notin S(2)$, define $p_{2n} \equiv \pi_2^*\alpha_n^{2*}$ and $q_{2n} = 0$. If $n\in S(2)$ but $n\notin S(1)$, define $p_{1n} \equiv \pi_1^*\alpha_n^{1*}$ and $q_{1n} = 0$. Thus we have complete overlapping price and quantity data for the two periods. Use the actual data along with the imputed data to calculate a favourite price index and define the companion quantity index residually by deflating the value ratio by the price index.

- A product or model with characteristics vector $z^* \equiv [z_1^*,...,z_K^*]$ should have a log price which is approximately equal to $\chi^{1*} + \Sigma_{k=1}^K \gamma_k^{1*} z_k^* \equiv \ln p^{1*}$ in period 1 and a log price which is approximately equal to $\chi^{2*} + \Sigma_{k=1}^K \gamma_k^{2*} z_k^* \equiv \ln p^{2*}$ in period 2. Choose z^* to be a characteristics vector that is *representative* for the set of products that exist in periods 1 and 2. Then the exponential of $\ln(p^{2*}/p^{1*}) = \chi^{2*} - \chi^{1*} + \Sigma_{k=1}^K (\gamma_k^{2*} - \gamma_k^{1*}) z_k^*$ can serve as a measure of average inflation over the period. The problem with this method is that there are many possible choices for the reference vector z^*.[122]
- Use each set of quality adjustment factors to generate two consistent measures of inflation over the two periods and then take the average of the two measures.

In what follows, we will work out the algebra for the third alternative.[123] Let δ^{1*} be the share weighted average of the quality adjusted log prices for period 1, p_{1n}/α_n^{2*}, using the period 2 quality adjustment factors α_n^{2*} defined by (184) and let δ^{2*} be the share weighted average of the quality adjusted log prices for period 2, p_{2n}/α_n^{1*}, using the period 1 quality adjustment factors α_n^{1*} defined by (184):

$$(187)\ \delta^{1*} \equiv \Sigma_{n\in S(1)}\, s_{1n} \ln(p_{1n}/\alpha_n^{2*})\ ;\ \delta^{2*} \equiv \Sigma_{n\in S(2)}\, s_{2n} \ln(p_{2n}/\alpha_n^{1*})\ .$$

It can be seen that $\chi^{2*} - \delta^{1*}$ is a *constant quality measure* of overall log price change which uses the quality adjustment factors α_n^{2*} for period 2 to deflate prices in both periods. Similarly, $\delta^{2*} - \chi^{1*}$ is a *constant quality measure* of overall log price change which uses the quality adjustment factors α_n^{1*} for period 1 to deflate prices in both periods. It is natural to take the arithmetic mean of these two measures of constant quality log price change in order to obtain the following counterpart, ρ_2^{**}, to the adjacent period time dummy measure of constant quality log price change, ρ_2^* defined by (177) above.

$$\begin{aligned}
(188)\ \rho_2^{**} &\equiv \tfrac{1}{2}[\chi^{2*} - \delta^{1*}] + \tfrac{1}{2}[\delta^{2*} - \chi^{1*}] \\
&= \tfrac{1}{2}[\Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^{2*}) - \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{2*})] \\
&\quad + \tfrac{1}{2}[\Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^{1*}) - \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{1*})] && \text{using (185)-(187)} \\
&= \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \tfrac{1}{2}(\ln\alpha_n^{1*} + \ln\alpha_n^{2*})] \\
&\quad - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \tfrac{1}{2}(\ln\alpha_n^{1*} + \ln\alpha_n^{2*})] \\
&= \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K (\tfrac{1}{2}\gamma_k^{1*} + \tfrac{1}{2}\gamma_k^{2*}) z_{nk}] \\
&\quad - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K (\tfrac{1}{2}\gamma_k^{1*} + \tfrac{1}{2}\gamma_k^{2*}) z_{nk}] && \text{using (184).}
\end{aligned}$$

Using (174), (177) can be written as follows:

[122] Note that if γ^{1*} turns out to equal γ^{2*}, then $\ln(p^{2*}/p^{1*}) = \chi^{2*} - \chi^{1*}$ and $\alpha^{1*} = \alpha^{2*}$ and $\chi^{2*} - \chi^{1*}$ turns out to equal ρ_2^* defined by (177).

[123] The analysis which follows is due to Silver and Heravi (2007), Diewert, Heravi and Silver (2009) and de Haan (2009).

(189) $\rho_2^* = \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \Sigma_{k=1}^{K}\, \gamma_k^* z_{nk}] - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^{K}\, \gamma_k^* z_{nk}]$.

The time dummy hedonic regression model defined by the minimization problem (170) uses the hedonic coefficients, γ_k^* for k = 1,..,K to form the quality adjustment factors α_n^* for n = 1,...,N. The single period hedonic regressions are defined by the minimization problems defined by (178) and (179), which in turn generate the two sets of hedonic coefficients, the γ_k^{1*} and the γ_k^{2*} for k = 1,..,K. But in the end, these two sets of hedonic coefficients are averaged when the overall measure of log price change defined by ρ_2^{**} is calculated. Thus the only difference between ρ_2^* defined by (189) and ρ_2^{**} defined by (188) is that the average hedonic coefficients $½\gamma_k^{1*} + ½\gamma_k^{2*}$ are used in (188) while ρ_2^* uses the single set of coefficients γ_k^*. Thus (189) lets the single regression do the job of constructing a set of hedonic coefficients that covers both periods (an internal aggregation) while (188) averages the results of the single period regressions (an external aggregation).

Which approach is "better"? The hedonic imputation approach requires the estimation of 2 + 2K parameters while the adjacent period time dummy hedonic approach requires only 2 + K parameters. Thus if the number of price observations in the two periods is plentiful, then the hedonic imputation approach will fit the data better and thus will generally be the preferred approach. However, as the above algebra shows, in the end, the indexes which are generated by the two approaches may not be all that different. If the number of observations is small and K is relatively large, then the adjacent period time dummy approach will probably generate less volatile measures of price change and hence is probably the preferred approach.[124]

A problem with all of the hedonic regression models that we have considered thus far is that the underlying economic model is quite restrictive; i.e., the underlying exact model is $p_{tn} = \pi_t\alpha_n$ which implies that purchasers of the products have linear preferences over the N products under consideration.[125] Linear preferences mean that the quality adjusted products are perfect substitutes for each other. In the following two sections, we will consider economic models which relax this assumption of perfect substitutes.

---

124 "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." Zvi Griliches (1971; 7). Thus Griliches suggested the time dummy approach if the separate hedonic regressions led to substantial fluctuation in the parameter estimates.

Griliches (1971; 7) suggested the time dummy approach if the single equation hedonic regres

125 This criticism of hedonic regression models is similar to that of Hausman (2003; 32): "In the presence of the introduction of new goods and quality improvement of existing goods, both prices and quantities (or alternatively, prices and expenditures) must be used to calculate a correct cost of living index. Using only prices and ignoring information in quantity data will never allow for a correct estimate of a cost of living index in the presence of new goods and improvements in existing goods." However, if the fit of a hedonic regression model is good, then the hedonic regression model is justified and there is no need to move to a more complicated consumer demand framework.

13. Estimating Reservation Prices: The Case of CES Preferences

In this section, we will explain Feenstra's (1994) Constant Elasticity of Substitution (CES) methodology that he proposed to measure the benefits and costs to consumers due to the appearance of new products and the disappearance of existing products.[126]

The Feenstra methodology starts out by making the same assumptions as were made in section 2; i.e., it is assumed that purchasers of a group of N products collectively maximize the linearly homogeneous, concave and nondecreasing aggregator or utility function F(q) subject to a budget constraint. Given that purchasers face the positive vector of prices $p \equiv (p_1,...,p_N)$, the *unit cost function* c(p) that is dual to the utility function f is defined as the minimum cost of attaining the utility level that is equal to one:

(190) $c(p) \equiv \min_q \{F(q) \geq 1; q \geq 0_N\}$.

If the unit cost function c(p) is known, then using duality theory, it is possible to recover the underlying utility function f(q).[127] Feenstra assumed that the unit cost function has the following *CES functional form*:

(191) $c(p) \equiv \alpha_0 [\sum_{n=1}^N \alpha_n p_n^{1-\sigma}]^{1/(1-\sigma)}$ if $\sigma \neq 1$;
$\equiv \alpha_0 \prod_{n=1}^N p_n^{\alpha_n}$ if $\sigma = 1$

where the α_i and σ are nonnegative parameters with $\sum_{i=1}^N \alpha_i = 1$. The unit cost function defined by (191) is a *Constant Elasticity of Substitution (CES) utility function* which was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961)[128].

The parameter σ is the *elasticity of substitution*;[129] when $\sigma = 0$, the unit cost function defined by (191) becomes linear in prices and hence corresponds to a fixed coefficients aggregator function which exhibits 0 substitutability between all commodities. When $\sigma = 1$, the corresponding aggregator or utility function is a Cobb-Douglas function. When σ approaches $+\infty$, the corresponding aggregator function f approaches a linear aggregator function which exhibits infinite substitutability between each pair of inputs. The CES unit cost function defined by (191) is of course *not* a fully flexible functional form (unless the number of commodities being aggregated is N = 2) but it is considerably more flexible

---

[126] The exposition in this section follows that of Diewert and Feenstra (2017).

[127] It can be shown that for $q >> 0_N$, $F(q) = 1/\max_p \{c(p): \Sigma_{n=1}^N p_n q_n \leq 1 ; p \geq 0_N\}$; see Diewert (1974; 110-112) on the duality between linearly homogeneous aggregator functions F(q) and unit cost functions c(p).

[128] In the mathematics literature, this aggregator function or utility function is known as a mean of order $r \equiv 1 - \sigma$; see Hardy, Littlewood and Polyá (1934; 12-13).

[129] Let c(p) be an arbitrary unit cost function that is twice continuously differentiable. The Allen (1938; 504) Uzawa (1962) *elasticity of substitution* $\sigma_{nk}(p)$ between products n and k is defined as $c(p)c_{nk}(p)/c_n(p)c_k(p)$ for $n \neq k$ where the first and second order partial derivatives of c(p) are defined as $c_n(p) \equiv \partial c(p)/\partial p_n$ and $c_{nk}(p) \equiv \partial^2 c(p)/\partial p_n \partial p_k$. For the CES unit cost function defined by (2), $\sigma_{nk}(p) = \sigma$ for all pairs of products; i.e., the elasticity of substitution between all pairs of products is a constant for the CES unit cost function.

than the zero substitutability aggregator function (this is the special case of (191) where σ is set equal to zero) or the linear aggregator function (which corresponds to $\sigma = +\infty$).

In order to simplify the notation, we set $r \equiv 1 - \sigma$. Under the assumption of cost minimizing behavior on the part of purchasers of the N products for periods t = 1,...,T, Shephard's (1953; 11) Lemma tells us that the observed period t consumption of commodity i, q_i^t, will be equal to $u^t\partial c(p^t)/\partial p_i$ where $\partial c(p^t)/\partial p_i$ is the first order partial derivative of the unit cost function with respect to the ith commodity price evaluated at the period t prices and $u^t = f(q^t)$ is the aggregate (unobservable) level of period t utility. As usual, denote the share of product i in total sales of the N products during period t as $s_{ti} \equiv p_{ti}q_{ti}/p^t\cdot q^t$ for i = 1,...,N and t = 1,...,T where $p^t\cdot q^t \equiv \sum_{n=1}^N p_{tn}q_{tn}$. We initially assume that there are no missing products. Note that the assumption of cost minimizing behavior during each period implies that the following equations will hold:

$$(192)\ p^t\cdot q^t = u^t c(p^t)\ ; \qquad t = 1,...,T$$

where c is the CES unit cost function defined by (191).

Using the CES functional form defined by (191) and assuming that $\sigma \neq 1$ (or $r \neq 0$),[130] the following equations are obtained using Shephard's Lemma:

$$(193)\ q_{ti} = u^t\alpha_0\,[\textstyle\sum_{n=1}^N \alpha_n\,(p_{tn})^r]^{(1/r)-1}\alpha_i\,(p_{ti})^{r-1}; \qquad i = 1,\ldots,N;\ t = 1,...,T$$
$$= u^t c(p^t)\ \alpha_i\,(p_{ti})^{r-1}/\textstyle\sum_{n=1}^N \alpha_n\,(p_{tn})^r\ .$$

Premultiply equation i for period t in (193) by $p_{ti}/p^t\cdot q^t$. Using (191) and (192), the resulting equations can be rewritten as follows:

$$(194)\ \ s_{ti} = \alpha_i\,(p_{ti})^r/\textstyle\sum_{n=1}^N \alpha_n\,(p_{tn})^r\ ; \qquad i = 1,\ldots,N;\ t = 1,...,T.$$

The NT share equations defined by (194) can be used as estimating equations using a nonlinear regression approach. Note that the positive scale parameter α_0 cannot be identified using equations (194), which of course is normal: utility can only be estimated up to an arbitrary scaling factor. Henceforth, we will assume $\alpha_0 = 1$. The share equations (194) are homogeneous of degree one in the parameters $\alpha_1,...,\alpha_N$ and thus the identifying restriction on these parameters, $\sum_{i=1}^N \alpha_i = 1$, can be replaced with an equivalent restriction such as $\alpha_N = 1$.

The sequence of *period t CES price indexes* (relative to the level of prices for period 1), P_{CES}^t, can be defined as the following ratios of unit costs in period t relative to period 1:

$$(195)\ P_{CES}^t \equiv [\textstyle\sum_{n=1}^N \alpha_n\,(p_{tn})^r]^{(1/r)} / [\textstyle\sum_{n=1}^N \alpha_n\,(p_{1n})^r]^{(1/r)}\ ; \qquad t = 1,...,T.$$

---

[130] When $\sigma = 1$, we have the case of Cobb-Douglas preferences. In the remainder of this paper, we will assume that $\sigma > 1$ (or equivalently, that $r < 0$).

Suppose further that the observed price and quantity data vectors, p^t and q^t for t = 1,...,T, satisfy equations (192) where c(p) is defined by (191) and the quantity data vectors q^t satisfy the Shephard's Lemma equations (193). This means that the observed price and quantity data are consistent with cost minimizing behavior on the part of purchasers where all purchasers have CES preferences that are dual to the CES unit cost function defined by (191). Then Sato (1976) and Vartia (1976) showed that the sequence of CES price indexes defined by (195) *could be numerically calculated just using the observed price and quantity data*; i.e., it is not necessary to estimate the unknown α_n and σ (or r) parameters in equations (195). The logarithm of the period t fixed base *Sato-Vartia Index* P_{SV}^t is defined by the following equation:

$$(196)\ \ln P_{SV}^t \equiv \Sigma_{n=1}^N w_n^t \ln(p_{tn}/p_{1n})\ ; \qquad t = 1,...,T.$$

The weights w_n^t that appear in equations (196) are calculated in two stages. The first stage set of weights is defined as $w_n^{t*} \equiv (s_{tn} - s_{1n})/(\ln s_{tn} - \ln s_{1n})$ for n = 1,...,N and t = 1,...,T provided that $s_{tn} \neq s_{1n}$. If $s_{tn} = s_{1n}$, then define $w_n^{t*} \equiv s_n^t = s_n^1$. The second stage weights are defined as $w_n^t \equiv w_n^{t*}/\Sigma_{i=1}^N w_i^{t*}$ for n = 1,...,N and t = 1,...,T. Note that in order for $\ln P_{CES}^t$ to be well defined, we require that $s_{tn} > 0$, $s_{1n} > 0$, $p_{tn} > 0$ and $p_{1n} > 0$ for all n = 1,...,N and t = 1,...,T; i.e., all prices and quantities must be positive for all products and for all periods.

With this background information in hand, we can explain Feenstra's (1994) model where "new" commodities can appear and "old" commodities can disappear from period to period.

Feenstra (1994) assumed CES preferences with $\sigma > 1$ (or equivalently, $r < 0$). He applied the reservation price methodology first introduced by Hicks (1940); i.e., as mentioned earlier, Hicks assumed that the consumer had preferences over all goods, but for the goods which had not yet appeared, there was a reservation price that would be just high enough that consumers would not want to purchase the good in the period under consideration.[131] This assumption works rather well with CES preferences, *because we do not have to estimate these reservation prices*; they will all be equal to $+\infty$ when $\sigma > 1$.

Feenstra allowed for new products to appear and for existing products to disappear from period to period.[132] Feenstra assumed that the set of commodities that are available in period t is S(t) for t = 1,...,T. The (imputed) prices for the unavailable commodities in each period are set equal to $+\infty$ and thus if $r < 0$, an infinite price p_{tn} raised to a negative power generates a 0; i.e., if product n is unavailable in period t, then $(p_{tn})^r = (\infty)^r = 0$ if r is negative.

[131] The same logic is applied to disappearing products.

[132] In many cases, a "new" product is not a genuinely new product; it is just a product that was not in stock in the previous period. Similarly, in many cases, a disappearing product is not necessarily a truly disappearing product; it is simple a product that was not in stock for the period under consideration. Many retail chains rotate products, temporarily discontinuing some products in favour of competing products in order to take advantage of manufacturer discounted prices for selected products.

The CES period t true price level under these conditions when $r < 0$ turns out to be the following CES unit cost function that is defined over only products that are available during period t:

$$(197)\ c(p^t) \equiv [\sum_{n=1}^{N} \alpha_n (p_{tn})^r]^{(1/r)} = [\sum_{n\in S(t)} \alpha_n (p_{1n})^r]^{1/r} .$$

Using equations (193) for this new model and multiplying the period t demand q_i^t by the corresponding price p_i^t for the items that are actually available leads to the following equations which describe the purchasers' nonzero expenditures on product i in period t:

$$(198)\ p_{ti}q_{ti} = u^t [\sum_{n\in S(t)} \alpha_n (p_{tn})^r]^{(1/r)-1}\alpha_i (p_{ti})^r ; \qquad t = 1,...,T;\ i\in S(t)$$
$$= u^t c(p^t)\ \alpha_i (p_{ti})^r/\sum_{n\in S(t)} \alpha_n (p_{tn})^r .$$

In each period t, the sum of observed expenditures, $\Sigma_{n\in S(t)}\ p_{tn}q_{tn}$, equals the period t utility level, u^t, times the CES unit cost $c(p^t)$ defined by (197):

$$(199)\ \Sigma_{n\in S(t)}\ p_{tn}\, q_{tn} = u^t c(p^t) = u^t[\sum_{i\in S(t)} \alpha_i (p_{1i})^r]^{1/r} ; \qquad t = 1,...,T.$$

Recall that the ith sales share of product i in period t was defined as $s_{ti} \equiv p_{ti}q_i^t/\Sigma_{n\in S(t)}\ p_{tn}q_{tn}$ for $t = 1,...,T$ and $i\in S(t)$. Using these share definitions and equations (199), we can rewrite equations (198) in the following form:

$$(200)\ s_{ti} = \alpha_i (p_{ti})^r/\sum_{n\in S(t)} \alpha_n (p_{tn})^r ; \qquad t = 1,...,T;\ i\in S(t)$$
$$= \alpha_i (p_{ti})^r/c(p^t)^r$$

where the second set of equations follows using definitions (197).

Now we can work out Feenstra's (1994) model for measuring the benefits and costs of new and disappearing commodities. Start out with the period t CES exact price level defined by (197) and define the CES fixed base price index for period t, P_{CES}^t, as the ratio of the period t CES price level to the corresponding period 1 price level:[133]

$$(201)\ P_{CES}^t \equiv c(p^t)/c(p^1) ; \qquad t = 1,...,T$$
$$= [\sum_{i\in S(t)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)} \alpha_i (p_{1i})^r]^{1/r}$$
$$= [\text{ Index 1}]\times[\text{Index 2}]\times[\text{Index 3}]$$

where the three indexes in equations (201) are defined as follows:

$$(202)\ \text{Index 1} \equiv [\sum_{i\in S(t)\cap S(1)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_i^1)^r]^{1/r} ;$$
$$(203)\ \text{Index 2} \equiv [\sum_{i\in S(t)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_{ti})^r]^{1/r} ;$$
$$(204)\ \text{Index 3} \equiv [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_{1i})^r]^{1/r} / [\sum_{i\in S(1)} \alpha_i (p_{1i})^r]^{1/r} .$$

---

[133] In the algebra which follows, the prices and quantities of period 1 can be replaced with the prices and quantities of any period. Feenstra (1994) developed his algebra for $c(p^t)/c(p^{t-1})$.

Note that Index 1 defines a CES price index over the set of commodities that are available in both periods t and 1. Denote the CES cost function c^{t*} that has the same α_n parameters as before but is now defined over only products that are available in periods 1 and t:

$$(205)\ c^{t*}(p) \equiv [\sum_{i\in S(t)\cap S(1)} \alpha_i (p_i)^r]^{1/r}\ ; \qquad t = 1,2,...,T.$$

The period t expenditure share equations that correspond to equations (200) using the unit cost functions defined by (205) are the following ones:

$$\begin{aligned}(206)\ s_i^{t*} &\equiv p_{ti}q_{ti}/\Sigma_{n\in S(t)\,)\cap S(1)}\ p_{tn}q_{tn} \qquad t = 1,...,T;\ i\in S(1)\cap S(t)\\ &= \alpha_i (p_{ti})^r/\sum_{n\in S(t)\,)\cap S(1)} \alpha_n (p_{tn})^r\\ &= \alpha_i (p_{ti})^r/c^{t*}(p^t)^r\end{aligned}$$

where the third equality follows using definitions (205).

Note that Index 1 is equal to $c^{t*}(p^t)/c^{t*}(p^1)$ and the Sato-Vartia formula (196) (restricted to commodities n that are present in periods 1 and t) can be used to calculate this index using the observed price and quantity data for the products that are available in both periods 1 and t.

We turn now to the evaluation of Indexes 2 and 3. It turns out that we will need an estimate for the elasticity of substitution σ (or equivalently of r) in order to find empirical expressions for these indexes.[134] It is convenient to define the following *observable expenditure or sales ratios*:

$$(207)\ \lambda^t \equiv \sum_{n\in S(t)} p_{tn}q_{tn}/\sum_{n\in S(1)\cap S(t)} p_{tn}q_{tn}\ ; \qquad t = 1,...,T;$$
$$(208)\ \mu^t \equiv \sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}\, q_{1n}\ ; \qquad t =1,...\ T.$$

We assume that there is at least one product that is present in periods 1 and t for each t. Let product i be any one of these common products for a given t. Then the share equations (200) and (206) hold for this product. These share equations can be rearranged to give us the following two equations:

$$(209)\ \alpha_i(p_{ti})^r = [\sum_{n\in S(t)} \alpha_n (p_{tn})^r]\ p_{ti}q_{ti}/[\sum_{n\in S(t)} p_{tn}q_{tn}]\ ;$$
$$(210)\ \alpha_i(p_{ti})^r = [\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{tn})^r]\ p_{ti}q_{ti}/[\sum_{n\in S(1)\cap S(t)} p_{tn}q_{tn}].$$

Equating (209) to (210) leads to the following equations:

$$\begin{aligned}(211)\ \sum_{n\in S(t)} \alpha_n (p_{tn})^r/\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{tn})^r &= \sum_{n\in S(t)} p_{tn}q_{tn}/\sum_{n\in S(1)\cap IS(t)} p_{tn}q_{tn}\\ &= \lambda^t\end{aligned}$$

where the last equality follows using definition (207). Now take the 1/r root of both sides of (211) and use definition (203) in order to obtain the following equality:

[134] See Diewert and Feenstra (2017) for a variety of methods for estimating the elasticity of substitution.

(2012) Index 2 = $[\lambda^t]^{1/r} = [\sum_{i\in S(t)} p_{ti}\, q_{ti}/\sum_{i\in S(1)\cap S(t)} p_{ti}q_{ti}]^{1/r}$.[135]

Again assume that product i is available in periods 1 and t. Rearrange the share equations (200) and (206) for t = 1 and product i and we obtain the following two equations:

(213) $\alpha_i(p_{1i})^r = [\sum_{n\in I(1)} \alpha_n (p_{1n})^r]\, p_{1i}q_{1i}/[\sum_{n\in S(1)} p_{1n}q_{1n}]$;
(214) $\alpha_i(p_{1i})^r = [\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{1n})^r]\, p_{1i}q_{1i}/[\sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}]$.

Equating (213) to (214) leads to the following equations:

(215) $\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{1n})^r/\sum_{n\in S(1)} \alpha_n (p_{1n})^r = \sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}q_{1n}$
$= \mu^t$

where the last equality follows using definition (208). Now take the 1/r root of both sides of (215) and use definition (201) in order to obtain the following equality:[136]

(216) Index 3 = $[\mu^t]^{1/r} = [\sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}q_{1n}]^{1/r}$.

Thus if r is known or has been estimated, then Index 2 and Index 3 can readily be calculated as simple ratios of sums of observable expenditures raised to the power 1/r. Note that $[\sum_{i\in S(t)} p_{ti}q_{ti}/\sum_{i\in S(1)\cap S(t)} p_{ti}q_{ti}] \geq 1$. If period t has products that were not available in period 1, then the strict inequality will hold and since $1/r < 0$, it can be seen that Index 2 will be less than unity. Thus Index 2 is a measure of how much the true cost of living index is *reduced* in period t due to the introduction of products that were not available in period 1. Similarly, $[\sum_{i\in I(1)\cap I(t)} p_i^1q_i^1/\sum_{i\in I(1)} p_i^1q_i^1] \leq 1$. If period 1 has products that are not available in period t, then the strict inequality will hold and since1/r < 0, it can be seen that Index 3 will be greater than unity, Thus Index 3 is a measure of how much the true cost of living index is *increased* in period t due to the disappearance of products that were available in period 1 but are not available in period t.

---

135 If new products become available in period t that were not available in period 1, then $\lambda^t > 1$. Recall that $r = 1 - \sigma$ and $r < 0$. Index 2 evaluated at period t prices equals $(\lambda^t)^{1/r} = (\lambda^t)^{1/(1-\sigma)}$ and thus is an increasing function of σ for $1 < \sigma < +\infty$. With $\lambda^t > 1$, the limit of $(\lambda^t)^{1/(1-\sigma)}$ as σ approaches 1 is 0 and the limit of $(\lambda^t)^{1/(1-\sigma)}$ as σ approaches $+\infty$ is 1. Thus the gains in utility from increased product variety are huge if σ is slightly greater than 1 and diminish to no gains at all as σ becomes very large. Suppose that λ^t =1.05 and σ = 1.01, 1.1, 1.5, 2, 3, 5, 10 and 100. Then Index 2 will equal 0.0076, 0.614, 0.907, 0.952, 0.976, 0.988, 0.995 and 0.9995 respectively. Thus the gains from increased product variety are very sensitive to the estimate for the elasticity of substitution. The gains are gigantic if σ is close to 1.

136 If some products that were available in period 1 become unavailable in period t, then $\mu^t < 1$. Index 3 evaluated at period 1 prices equals $(\mu^t)^{1/r} = (\mu^t)^{1/(1-\sigma)}$ and is a decreasing function of σ for $1 < \sigma < +\infty$. With $\mu^t < 1$, the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches 1 is $+\infty$ and the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches $+\infty$ is 1. Thus the losses in utility from decreased product variety are huge if σ is slightly greater than 1 and diminish to no gains at all as σ becomes very large. Suppose that μ^t =0.95 and σ takes on the same values as in the previous footnote. Then Index 3 will equal 168.9, 1.670, 1.108, 1.053, 1.026, 1.013, 1.0057 and 1.00052 respectively. Thus the losses are gigantic if σ is close to 1 and negligible if σ is very large.

Turning briefly to the problems associated with estimating r (and the α_n) when not all products are available in all periods, it can be seen that the initial estimating share equations (191) are replaced by the following estimating equations:

$$(217)\ s_{tn} = \alpha_n (p_{tn})^r / \Sigma_{k\in S(t)}\, \alpha_k (p_{tk})^r\ ; \qquad t = 1,...,T;\ n\in S(t).$$

However, there are many methods that have been suggested in the literature to estimate r (or the elasticity of substitution σ) when there are missing products; see for example Diewert and Feenstra (2017).

The Feenstra methodology is easy to implement once an estimate for σ is available and so it has been widely used in the macroeconomic literature. However, if the elasticity of substitution is low and new products outnumber disappearing products, then this methodology will lead to quality adjusted price indexes which will decrease by amounts that are not plausible and this point should be kept in mind.[137] The Feenstra methodology will tend to be biased for elasticities of substitution which are close to one and should not be used in this case.[138] Thus in the next section, we will study a model which is similar to Feenstra's model but the reservation prices generated by the model are finite and a flexible functional form for Q(q) is used in place of the CES functional form.

14. Estimating Reservation Prices: The Case of KBF Preferences

The functional form for the aggregator function F(q) that we will use in this section is the *KBF function form*, $Q_{KBF}(q) \equiv [q\cdot Aq]^{1/2}$ defined by (17) in section 4.[139] The system of *inverse demand functions* for this functional form for our data set with no missing observations is given by the following system of equations:

$$(218)\ p^t = P^t\, \nabla_q Q_{KBF}(q^t) = P^t\, [q^t\cdot Aq^t]^{-1/2}\, Aq^t\ ; \qquad t = 1,...,T$$

where the N by N matrix $A \equiv [a_{nk}]$ is symmetric (so that $A^T = A$) and thus has N(N+1)/2 unknown a_{nk} elements. As in section 4,we also assume that A has one positive eigenvalue with a corresponding strictly positive eigenvector and the remaining N−1 eigenvalues are negative or zero. These conditions will ensure that the aggregator function has indifference surfaces with the correct curvature.

The period t aggregate quantity is $Q^t \equiv [q^t\cdot Aq^t]^{1/2}$ for t = 1,...,T. Multiply the right hand side of equation t in (218) by $1= Q^t/[q^t\cdot Aq^t]^{1/2}$ for t = 1,...,T and we obtain the following system of estimating equations:

---

[137] Also keep in mind that the Feenstra methodology does not work at all if the elasticity of substitution is equal to or less than one.

[138] Another feature of the Feenstra methodology is that the reservation prices are infinite. Typically it does not take an infinitely high price to deter consumers from buying the product under consideration.

[139] The analysis in this section follows that of Diewert and Feenstra (2017). The same theoretical framework was suggested by Diewert (1980; 498-503) but a different flexible functional form was used to illustrate the methodology. The Diewert and Feenstra functional form is a better choice since the correct curvature conditions can be imposed on the KBF functional form without destroying its flexibility.

(219) $p^t = P^tQ^tAq^t/q^t{\cdot}Aq^t = v_tAq^t/q^t{\cdot}Aq^t$; t = 1,...,T

where we have used equations (9), $P^tQ^t = p^t{\cdot}q^t = v_t$ for t = 1,...,T, to derive the second set of equations in (219). Now convert equations (219) into a set of share equations by taking component n in the vector p^t, p_{tn}, and multiplying both sides of this equation by q_{tn} and dividing by $v_t = p^t{\cdot}q^t$. We obtain the following system of estimating equations:

(220) $s_{tn} = \Sigma_{m=1}^{N} q_{tn}a_{nm}q_{tm}/\Sigma_{n=1}^{N} \Sigma_{m=1}^{N} q_{tm}a_{nm}q_{tm}$; t = 1,...,T; n = 1,...,N.

Now introduce missing products into the model. Let S(t) be the set of products n that are present in period t for t = 1,...,T. If product n is missing in period t, define $q_{tn} \equiv 0$ and s_{tn} = 0. Define q^t and s^t as the period t vectors of quantities and shares where $q_{tn} \equiv 0$ and $s_{tn} \equiv 0$ if product n is missing in period t. It can be seen that equations (220) are still valid when there are missing products except that instead of t = 1,...,T; n = 1,...,N, we have t = 1,...,T and $n \in S(t)$. Thus we use equation t,n in (220) as an estimating equation only if the corresponding product n is present in period t.

The N(N+1)/2 unknown parameters a_{nm} in the A matrix can be determined by solving the following nonlinear least squares minimization problem:[140]

(221) $\min_A \Sigma_{t=1}^{T} \Sigma_{n \in S(t)} [s_{tn} - \{\Sigma_{m=1}^{N} q_{tn}a_{nm}q_{tm}/\Sigma_{i=1}^{N} \Sigma_{j=1}^{N} q_{ti}a_{ij}q_{tj}\}]^2$.

Note that the minimization problem defined by (221) is run as a single nonlinear regression rather than as a system of N share equations, which is a more traditional approach when estimating systems of consumer demand functions. The unusual specification is due to the fact that there are missing products in the T time periods and so the traditional systems approach cannot be applied. A second point to note is that not all of the parameters a_{nm} can be identified: if a_{nm}^* solves (221), then so does λa_{nm}^* for $1 \leq n \leq m \leq N$ for all $\lambda \neq 0$. Thus a normalization on the matrix of parameters is required for a unique solution to (221). A final point to note is that the error terms in (221) are not weighted by their economic importance. There is no need to do this because the dependent variables in (221), the shares, are already weighted by their economic importance and so there is no need for further weighting. Put another way, each share is equally important (and is measured in comparable units) and hence it makes sense to fit the observed shares by model predicted shares using a least squares approach.

Once the parameters a_{nm}^* have been determined, we can use the price equations defined by (219) above to determine the *Hicksian reservation prices* p_{tn}^* for the missing products for t = 1,...,T and n does not belong to S(t):

(222) $p_{tn}^* \equiv v_t\{\Sigma_{m=1}^{N} a_{nm}^*q_{tm}/\Sigma_{i=1}^{N} \Sigma_{j=1}^{N} q_{ti}a_{ij}^*q_{tj}\}$; t = 1,...,T; $n \notin S(t)$.

[140] Alternative estimating equations are considered in Diewert and Feenstra (2017).

Note that the reservation prices defined by (222) will be finite. Using the observed prices and quantities for each period t along with the imputed prices p_{tn}^*, complete price and quantity vectors for each period can be formed. These complete price and quantity vectors can be used to form price and quantity levels for each period using a preferred index number formula. Alternatively, the estimated parameters a_{nm}^* can be used to form the matrix of parameters, $A^* \equiv [a_{nm}^*]$. Use the estimated A^* matrix to form the period t quantity levels, $Q^{t*} \equiv [q^t \cdot A^* q^t]^{1/2}$ for t = 1,...,T and the corresponding period t price levels, $P^{t*} \equiv v_t/Q^{t*}$ for t = 1,...,T.

There are two problems with the above methodology that need to be addressed: (i) how can we be sure that the estimated A matrix satisfies the eigenvalue restrictions listed above and (ii) how can we estimate the parameters of the A matrix when N is large?

The number of unknown parameters in the A matrix is N(N+1)/2 if there are N products in the window of observations. If N = 10, N(N+1)/2 = 55; if N = 100, N(N+1)/2 = 5050. Thus it will be impossible to estimate all of the parameters in the A matrix if N is large.

The above two difficulties with this methodology can be addressed if we make use of the following reparameterization of the A matrix. Thus we set A equal to the following expression:[141]

(223) $A = bb^T + B$; $b >> 0_N$; $B = B^T$; B is negative semidefinite; $Bq^* = 0_N$.

The vector $b^T \equiv [b_1,...,b_N]$ is a row vector of positive constants and so bb^T is a rank one positive semidefinite N by N matrix. The symmetric matrix B has N(N+1)/2 independent elements b_{nk} but the N constraints $Bq^* = 0_N$ reduce this number by N. Thus there are N independent parameters in the b vector and N(N−1)/2 independent parameters in the B matrix so that $bb^T + B$ has the same number of independent parameters as the A matrix.

The reparameterization of A by $bb^T + B$ is useful in the present context because this reparameterization can be used to estimate the unknown parameters in stages. Thus initially set $B = O_{N\times N}$, a matrix of 0's. The resulting aggregator function becomes $F(q) = (q^T bb^T q)^{1/2} = (b^T q b^T q)^{1/2} = b^T q$, a *linear utility function*. Thus this special case of (223) boils down to the linear utility function model that has been used repeatedly in this paper.

The matrix B is required to be negative semidefinite. The procedure used by Wiley, Schmidt and Bramble (1973) and Diewert and Wales (1987) can be used to impose negative semidefiniteness on B by setting B equal to $-CC^T$ where C is a lower triangular matrix.[142] Write C as $[c^1,c^2,...,c^N]$ where c^k is a column vector for k = 1,...,K. If C is lower triangular, then the first k−1 elements of c^k are equal to 0 for k = 2,3,...,N. The following representation for B will hold:

---

[141] Notation: b is regarded as a column vector and b^T is its transpose, which is a row vector.

[142] $C = [c_{nk}]$ is a lower triangular matrix if $c_{nk} = 0$ for k > n; i.e., there are 0's in the upper triangle. Wiley, Schmidt and Bramble showed that setting $B = -CC^T$ where C was lower triangular was sufficient to impose negative semidefiniteness while Diewert and Wales showed that any negative semidefinite matrix could be represented in this fashion.

$$(224)\ B = -\,CC^T = -\,\Sigma_{n=1}^{N} c^n c^{nT}$$

where the following restrictions on the vectors c^n are imposed in order to impose the restrictions $Bq^* = 0_N$ on B:[143]

$$(225)\ c^n \cdot q^* = 0\ ; \qquad n = 1,....,N.$$

As mentioned above, if N is not small, then typically, it will not be possible to estimate all of the parameters in the C matrix. Furthermore, typically nonlinear estimation is not successful if one attempts to estimate all of the parameters at once. Thus it is necessary to estimate the parameters in the utility function $F(q) = (q^TAq)^{1/2}$ in stages. In the first stage, estimate the linear utility function $F(q) = b^Tq$.[144] In the second stage, estimate $F(q) = (q^T[bb^T - c^1c^{1T}]q)^{1/2}$ where $c^{1T} \equiv [c_1{}^1,c_2{}^1,...,c_N{}^1]$ and $c^{1T}q^* = 0$. For starting coefficient values in the second nonlinear regression, use the final estimates for b from the first nonlinear regression and set the starting $c^1 \equiv 0_N$.[145] In the third stage, estimate $F(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T}]q)^{1/2}$ where $c^{1T} \equiv [c_1{}^1,c_2{}^1,...,c_N{}^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2{}^1,...,c_N{}^1]$ and $c^{2T}q^* = 0$. The starting coefficient values are the final values from the second stage with $c^2 \equiv 0_N$. At each stage, the log likelihood will generally increase.[146] Stop adding columns to the C matrix when the increase in the log likelihood becomes small (or the number of degrees of freedom becomes small). At stage k of this procedure, it turns out that a substitution matrix of rank k−1 is estimated that is the most negative semidefinite that the data will support.[147] This is the same type of procedure that Diewert and Wales (1988) used in order to estimate normalized quadratic preferences and they termed the final functional form a *semiflexible functional form*. The above treatment of the KBF functional form also generates a semiflexible functional form.

The above functional form for the aggregator function is more general than the linear utility function that has been used throughout most of this paper and it is more general than the CES aggregator function that was used in the previous section. Moreover, the reservation prices that the method generates are finite. Finally, the present model can deal with situations where a new product has a low elasticity of substitution with all existing products; i.e., it provides a more satisfactory solution to the new good problem. However, it has the drawback of being rather complex and hence it may be resistant to large scale applications of the method.

---

143 The restriction that C be upper triangular means that c^N will have at most one nonzero element, namely $c_N{}^N$. However, the positivity of q^* and the restriction $c^{NT}q^* = 0$ will imply that $c^N = 0_N$. Thus the maximal rank of B is N−1.

144 In order to identify all of the parameters, set one component of the b vector to equal 1.

145 We also use the constraint $c^{1T}q^* = 0$ to eliminate one of the $c_n{}^1$ from the nonlinear regression.

146 If it does not increase, then the data do not support the estimation of a higher rank substitution matrix and we stop adding columns to the C matrix. The log likelihood cannot decrease since the successive models are nested.

147 For a worked example of this methodology, see Diewert and Feenstra (2017).

This completes our selective review of quality adjustment methods that are based on economic approach to index numbers applied to purchasers of consumer goods and services.

15. Conclusion

The paper has taken a consumer demand perspective to addressing the problem of adjusting price and quantity indexes to take into account the benefits and costs of the introduction of new goods and services and the disappearance of existing commodities. This perspective allows all of the major methods that address the new and disappearing goods problem to be compared in a common framework.

There are three main methods that have been suggested in the literature to address the new goods problem: (i) the use of inflation adjusted carry forward and backward prices; (ii) hedonic regression methods and (iii) the estimation of consumer preferences and Hicksian reservation prices using both price and quantity data. The first two methods will work well if the new and disappearing products are highly substitutable with continuing products. However, if substitution is low, then the use of the first two methods will lead to substantial biases in price and quantity indexes for the class of products under consideration. In the low elasticity of substitution case, the third class of methods should be used; i.e., one should use either the cost or expenditure function methods suggested by Hausman[148] or the direct utility function estimation methods suggested by Diewert and Feenstra in section 14 above. Unfortunately, these methods are not easy to implement. Thus more research on these methods is required before statistical agencies can implement these methods on large scale.

Some of the more important points made in the paper are summarized below.

- Using the theoretical framework explained in section 2 and applying it to hedonic regressions in section 5 (when price and quantity data are available) shows that the hedonic regression approach generates two distinct estimates for the resulting price and quantity levels generated by the regression (unless the regression fits the data perfectly, in which case the two methods generate identical estimates). Thus

[148] "Ultimately, data on price and product attributes alone will not allow correct estimation of the compensating variation adjustment to a cost of living index. Quantity data are also needed, so that estimates of the demand functions (or equivalently, the expenditure or utility functions) can occur. For this reason, I disagree with the panel's conclusion that hedonic methods are 'probably the best hope' for improving quality adjustments (Schultze and Mackie (2002; 64 and 122)) since hedonic methods do not use quantity data to estimate consumer valuation of a product, and consumer demand must be the basis of a cost of living index." Jerry Hausman (2003; 37). We agree with Hausman's criticisms of hedonic regression techniques to deal with the quality change problem except that we note that hedonic regressions can work well if the class of products under consideration are close substitutes for each other. Also in some situations, we have no choice but to work with hedonic regressions rather than estimate consumer demand systems. For example, when constructing property price indexes, each property is a *unique good*, both over time and space. A property has a unique location and over time the structure on the property changes due to renovations and depreciation. Thus hedonic regressions must be used in this situation.

statistical agencies will have to choose between these two alternative index number estimates.

- The use of weights that reflect economic importance is recommended when running hedonic regressions; see the summary of the work of de Haan and Krsinich (2018) in section 11.
- In the two period context, section 7 shows how the use of weights can transform the problematic price index that results from an unweighted time product dummy hedonic regression into a superlative bilateral price index.
- The usefulness of the adjacent period weighted time product dummy hedonic regressions studied in section 8 is questionable; i.e., it may be preferable to use the carry forward and backward technique explained in section 4 in the bilateral case.
- Section 12 deals with hedonic regressions in the context of taste change. The results in this section indicate that the apparent increased flexibility offered by running separate hedonic regressions for each period is tempered by the need to average the results of the separate regressions in order to obtain proper price indexes. In the end, the use of a time dummy approach is recommended if the number of parameters is large relative to the number of observed prices.
- Hedonic regression models viewed from the Hicksian approach to the treatment of new products have a fundamental problem: the underlying economic model assumes that the products are perfect substitutes after the implied quality adjustment. This is not a problem if in fact, the quality adjusted products are close to being perfect substitutes but it can be a problem if this is not the case.
- The CES methodology for accounting for the benefits of new products due to Feenstra explained in section 13 can work well if the elasticity of substitution between the products under consideration is high. If it is not high, the method will tend to lead to price indexes which have a downward bias.
- The econometric method for dealing with new and disappearing products in the context of the Hicksian reservation price methodology avoids the problems associated with the Feenstra methodology but at the cost of a great deal of econometric complexity. A robust simplified version of this methodology is required before it can be applied by statistical agencies on a routine basis.

This paper has taken an economic approach to the problem of quality adjustment that is based on the basic model of household behavior explained in section 2. This economic model is not without its problems but it does lead to a unified approach to the treatment of quality change from an economic perspective.

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.

Aizcorbe, A. (2014), *A Practical Guide to Price Index and Hedonic Techniques*, Oxford, UK: Oxford University Press.

Allen, R.G.C. (1938), Mathematical Analysis for Economists, London: Macmillan.

Arrow, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow (1961), "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics* 63, 225-250.

Australian Bureau of Statistics (2016), "Making Greater Use of Transactions Data to Compile the Consumer Price Index", Information Paper 6401.0.60.003, November 29, Canberra: ABS.

Balk, B.M. (1980), "A Method for Constructing Price Indices for Seasonal Commodities", *Journal of the Royal Statistical Society, Series A* 143, 68-75).

Court, A.T. (1939), "Hedonic Price Indexes with Automotive Examples", pp. 99-117 in *The Dynamics of Automobile Demand*, New York: General Motors Corporation.

Dalén, J. (2001), "Statistical Targets for Price Indexes in Dynamic Universes," Paper presented at the sixth meeting of the Ottawa Group, April 2-6, Canberra, 2001.

Diewert, W.E., (1974), "Applications of Duality Theory," pp. 106-171 in M.D. Intriligator and D.A. Kendrick (ed.), *Frontiers of Quantitative Economics*, Vol. II, Amsterdam: North-Holland.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", *Econometrica* 46, 883-900.

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

Diewert, W.E. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited", *Journal of Productivity Analysis* 3, 211-248.

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. (2002), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", Department of Economics, Discussion Paper 02-15, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Diewert, W.E. (2003a), “Hedonic Regressions: A Consumer Theory Approach”, in Scanner Data and Price Indexes, Studies in Income and Wealth (Vol. 61), eds. R.C. Feenstra and M.D. Shapiro, Chicago: University of Chicago Press, pp. 317-348.

Diewert, W.E. (2003b), “Hedonic Regressions: A Review of Some Unresolved Issues”, Paper presented at the Seventh Meeting of the Ottawa Group, Paris, 27–29 May.

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.

Diewert, W.E. (2005a), “Weighted Country Product Dummy Variable Regressions and Index Number Formulae”, *Review of Income and Wealth* 51, 561-570.

Diewert, W.E. (2005b), “Adjacent Period Dummy Variable Hedonic Regressions and Bilateral Index Number Theory”, *Annales D’Économie et de Statistique,* No. 79/80, 759-786.

Diewert, W.E. (2018), “Scanner Data, Elementary Price Indexes and the Chain Drift Problem”, Discussion Paper 18-06, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

Diewert, W.E. and R. Feenstra (2017), “Estimating the Benefits and Costs of New and Disappearing Products”, Vancouver School of Economics, Discussion Paper 17-10, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

Diewert, W.E. and K.J. Fox (2017), “Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data”, Discussion Paper 17-02, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

Diewert, W.E., K.J. Fox and P. Schreyer (2017), “The Digital Economy, New Products and Consumer Welfare”, Vancouver School of Economics Discussion Paper 17-09, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

Diewert, W.E., J. de Haan and R. Hendriks (2015), “Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components”, *Econometric Reviews* 34, 106-126.

Diewert, W.E., S. Heravi and M. Silver (2009), “Hedonic Imputation versus Time Dummy Hedonic Indexes”, pp. 87-116 in W.E. Diewert, J. Greenlees and C. Hulten (eds.), *Price Index Concepts and Measurement*, Studies in Income and Wealth, Vol. 70. Chicago: University of Chicago Press.

Diewert, W.E. and R.J. Hill (2010), “Alternative Approaches to Index Number Theory”, pp. 263-278 in *Price and Productivity Measurement,* W.E. Diewert, Bert M. Balk, Dennis Fixler, Kevin J. Fox and Alice O. Nakamura (eds.), Victoria Canada: Trafford Press.

Diewert, W.E., N. Huang and K. Burnett-Issacs (2017), "Alternative Approaches for Resale Housing Price Indexes", Discussion Paper 17-05,Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1L4.

Diewert, W. E. and C. Shimizu (2016), "Hedonic Regression Models for Tokyo Condominium Sales," *Regional Science and Urban Economics* 60, 300-315.

Diewert, W. E. and C. Shimizu (2017), "Alternative Approaches to Commercial Property Price Indexes for Tokyo," *Review of Income and Wealth* 63:3, 492-519.

Diewert, W.E. and T.J. Wales (1987), "Flexible Functional Forms and Global Curvature Conditions", *Econometrica* 55, 43-68.

Diewert, W.E. and T.J. Wales (1988), "A Normalized Quadratic Semiflexible Functional Form", *Journal of Econometrics* 37, 327-42.

Eurostat (2013), *Handbook on Residential Property Prices Indices (RPPIs)*, Luxembourg: Publications Office of the European Union.

Feenstra, R.C. (1994), "New Product Varieties and the Measurement of International Prices", *American Economic Review* 84:1, 157-177.

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

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

Gorajek, A. (2018), "Econometric Perspectives on Economic Measurement", Research Discussion Paper 2018-08, Reserve Bank of Australia, 65 Martin Pl, Sydney NSW 2000.

de Haan, J. (2004a), "The Time Dummy Index as a Special Case of the Imputation Törnqvist Index," paper presented at The Eighth Meeting of the International Working Group on Price Indices (the Ottawa Group), Helsinki, Finland.

de Haan, J. (2004b), "Estimating Quality-Adjusted Unit Value Indices: Evidence from Scanner Data," Paper presented at the Seventh EMG Workshop, Sydney, Australia, December 12–14.

de Haan, J. (2009), "Comment on "Hedonic Imputation Versus Time Dummy Hedonic Indexes", pp 196-200 in W.E. Diewert, J. Greenlees, and C. Hulten (eds.), *Price Index Concepts and Measurements*, Chicago: University of Chicago Press.

de Haan, J. (2010), "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-pricing Methods", *Jahrbücher für Nationökonomie und Statistik* 230, 772-791.

de Haan, J. (2015), "A Framework for Large Scale Use of Scanner Data in the Dutch CPI", Paper presented at the 14th meeting of the Ottawa Group, May 22, Tokyo.

de Haan, J. and F. Krsinich (2012), "The Treatment of Unmatched Items in Rolling Year GEKS Price Indexes: Evidence from New Zealand Scanner Data", paper presented at the Economic Measurement Group Workshop 2012, Australian School of Business, University of New South Wales, November 23.

de Haan, J. and F. Krsinich (2014), "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes", *Journal of Business and Economic Statistics* 32:3, 341-358.

de Haan, J. and F. Krsinich (2018), "Time Dummy Hedonic and Quality Adjusted Unit Value Indexes: Do They Really Differ?", *Review of Income and Wealth* 64:4, 757-776.

Hardy, G.H., J.E. Littlewood and G. Polyá (1934), *Inequalities*, Cambridge: Cambridge University Press.

Hausman, J.A. (1981), "Exact Consumer's Surplus and Deadweight Loss", *American Economic Review* 71:4, 662-676.

Hausman, J.A. (1996), "Valuation of New Goods under Perfect and Imperfect Competition", pp. 20 -236 in *The Economics of New Goods*, T.F. Bresnahan and R.J. Gordon (eds.), Chicago: University of Chicago Press.

Hausman, J.A. (1999), "Cellular Telephone, New Products and the CPI", *Journal of Business and Economic Statistics* 17:2, 188-194.

Hausman, J. (2003), "Sources of Bias and Solutions to Bias in the Consumer Price Index", *Journal of Economic Perspectives* 17:1, 23-44.

Hicks, J.R. (1940), "The Valuation of the Social Income", *Economica* 7, 105–124.

Hill, R.J. (2013), "Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy", *Journal of Economic Surveys* 27, 879-914.

*Hulten, C. (2003), "Price Hedonics: A Critical Review", *Economic Policy Review* 9:3, 5-15.

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

Ivancic, L., W.E. Diewert and K.J. Fox (2009), "Scanner Data, Time Aggregation and the Construction of Price Indexes", Discussion Paper 09-09, Department of Economics, University of British Columbia, Vancouver, Canada.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Jevons, W.S., (1863), "A Serious Fall in the Value of Gold Ascertained and its Social Effects Set Forth (1863)", reprinted as pp. 13-118 in *Investigations in Currency and Finance* (1884), London: Macmillan and Co.

Jevons, W.S., (1865), "The Variation of Prices and the Value of the Currency since 1782", *Journal of the Statistical Society of London* 28, 294-320.

Keynes, J.M. (1909), "The Method of Index Numbers with Special Reference to the Measurement of General Exchange Value", reprinted as pp. 49-156 in *The Collected Writings of John Maynard Keynes* (1983), Volume 11, D. Moggridge (ed.), Cambridge: Cambridge University Press.

Keynes, J.M. (1930), *Treatise on Money*, Vol. 1. London: Macmillan.

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

Konüs, A.A. and S.S. Byushgens (1926), "K Probleme Pokupatelnoi Cili Deneg", *Voprosi Konyunkturi* 2, 151-172.

Krsinich, F. (2016), "The FEWS Index: Fixed Effects with a Window Splice', *Journal of Official Statistics* 32, 375-404.

Lehr, J. (1885), *Beiträge zur Statistik der Preise*, Frankfurt: J.D. Sauerländer.

Marshall, A. (1887), "Remedies for Fluctuations of General Prices', *Contemporary Review* 51, 355-375.

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

Pakes, A. (2001), "A Reconsideration of Hedonic Price Indices with and Application to PCs", NBER Working Paper 8715, Cambridge MA..

Rao, D.S. Prasada (1995), "On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons", Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D.S. Prasada (2004), "The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power parities in the ICP", paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30-July 3, 2004, Vancouver, Canada.

Rao, D.S. Prasada (2005), "On the Equivalence of the Weighted Country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons", Review of Income and Wealth 51:4, 571-580.

Rao, D.S. Prasada and G. Hajargasht (2016), "Stochastic Approach to Computation of Purchasing Power Parities in the International Comparison Program", *Journal of Econometrics* 191:2, 414-425.

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

Samuelson, P.A. and S. Swamy (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis", *American Economic Review* 64, 566-593.

Sato, K. (1976), "The Ideal Log-Change Index Number", *Review of Economics and Statistics* 58, 223-228.

Schultze, C. and C. Mackie (eds.), *At What Price?*, Washington DC: National Academy Press.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton: Princeton University Press.

Shimizu, C., K. Nishimura and T. Watanabe (2010), "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures", *Journal of Economics and Statistics* 230, 792-813.

Silver, M. (2010), "The Wrong and Rights of Unit Value Indices", *Review of Income and Wealth* 56:S1, 206-206.

Silver, M. (2011), "An Index Number Formula Problem: the Aggregation of Broadly Comparable Items, *Journal of Official Statistics* 27:4, 1–17.

Silver, M.S. (2018), "How to Measure Hedonic Property Price Indexes Better", *EURONA* 1/2018, 35-66.

Silver, M. and S. Heravi (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, 269-281.

Silver, M. and S. Heravi (2007), “The Difference Between Hedonic Imputation Indexes and Time Dummy Hedonic Indexes”, *Journal of Business and Economic Statistics* 25, 239-246.

Summers, R. (1973), “International Price Comparisons Based Upon Incomplete Data”, Review of Income and Wealth 19, 1-16.

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

Törnqvist, L. (1936), “The Bank of Finland's Consumption Price Index”, *Bank of Finland Monthly Bulletin* 10, 1-8.

Törnqvist, L. and E. Törnqvist (1937), Vilket är förhällandet mellan finska markens och svenska kronans köpkraft?”, *Ekonomiska Samfundets Tidskrift* 39, 1-39 reprinted as pp. 121-160 in *Collected Scientific Papers of Leo Törnqvist*, Helsinki: The Research Institute of the Finnish Economy, 1981.

Triplett, J. (1987), “Hedonic Functions and Hedonic Indexes”, pp. 630-634 in: John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Volume 2. New York, NY: Stockton Press.

Triplett, J. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, Directorate for Science, Technology and Industry, DSTI/DOC(2004)9, Paris: OECD.

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

Uzawa, H. (1962), “Production Functions with Constant Elasticities of Substitution”, *Review of Economic Studies* 29, 291-299.

Vartia, Y.O. (1976), “Ideal Log-Change Index Numbers”, *Scandinavian Journal of Statistics* 3, 121-126.

von Auer, L. (2014), “The Generalized Unit Value Index Family”, *Review of Income and Wealth* 60, 843-861.

von Hofsten, E. (1952), *Price Indexes and Quality Change*, London: George Allen and Unwin.

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

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

Walsh, C.M. (1921b), “The Best Form of Index Number: Discussion,” *Quarterly Publication of the American Statistical Association* 17, 537–544.

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

Wiley, D.E., W.H. Schmidt and W.J. Bramble (1973), “Studies of a Class of Covariance Structure Models”, *Journal of the American Statistical Association* 68, 317-323.

Wold, H. (1944), “A Synthesis of Pure Demand Analysis, Part 3”, *Skandinavisk Aktuarietidskrift* 27, 69-120.