Quality Adjustment Methods

April 29, 2020

Erwin Diewert,[1]
Discussion Paper 20-08,
Vancouver School of Economics,
The University of British Columbia,
Vancouver, Canada V6T 1L4.

Abstract: The paper attempts to present a unified approach to the problem of adjusting products for changes in quality using a consumer perspective. The basic assumption is that purchasers of a group of related products have identical homothetic preferences defined over the products and they maximize utility subject to expenditure constraints. Inflation adjusted carry forward or backward price adjustment can be put into this framework as can hedonic regressions and the estimation of Hicksian reservation prices for the missing products. Feenstra's Constant Elasticity of Substitution (CES) new products methodology also is a special case of the general methodology.

Key Words: Quality change adjustment, hedonic regressions, Hicksian reservation prices, consumer theory, inflation adjusted carry forward prices, econometric estimation of preferences, new and disappearing goods.

Journal of Economic Literature Classification Numbers: C43, C13, C32, D11, E31, I31

[1] University of British Columbia and University of New South Wales. Email: erwin.diewert@ubc.ca . This is a draft of Chapter 8 of *Consumer Price Index Theory*. The author thanks Jan de Haan, Adam Gorajek, Ronald Johnson and Chihiro Shimizu for helpful comments and Jerry Hausman and Mick Silver for helpful discussions over the years. This chapter is largely based on Diewert (2019).

1. Introduction

This chapter 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 chapter 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 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

[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 was explained in Chapter 5 and will be summarized 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).
[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

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 chapter.[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 $f(q)$ 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 $\max_q \{f(q) : p\cdot q = v\}$. Then given a functional form for f, the solution q to the utility maximization problem will satisfy the usual consumer demand functions, $q_n = d_n(p,v)$ for $n = 1,...,N$ where $d_n(p,v)$ 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 simplify the derivation of the consumer demand functions.[8] This is the approach used by Hausman (1981) (1996) (1999) (2003) to estimate reservation prices. However, the econometrics of this method are complex. To illustrate these problems, suppose that in the first 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 $0 = d_1(p_1^{1*},p_2^1,...,p_N^1,v^1) + e_1^1$ where $d_1(p,v)$ 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.

---

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.

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

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 sections 9 and 10 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 $f(q) = \Sigma_{n=1}^N \alpha_n q_n \equiv \alpha \cdot q$ 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 5-8. Furthermore, it turns out that the assumption of an underlying linear utility function provides a rationale for *hedonic regression models*, which will be studied in sections 5-8 below.

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. 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 can be replaced by $\ln p_{tn} = \ln\pi_t + \ln\alpha_n + e_{tn}$ for n = 1,...,N and t = 1,...,T where the e_{tn} are error terms. This is a stochastic model which was discussed in Chapter 7. It is also a special case of a *hedonic regression model* where prices are regressed on the characteristics of products. In this simple framework, each product has its own separate characteristic. Estimators for the logarithms of π_t and α_n are found by

[9] Suppose that the utility function f(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, $p_n = [\partial f(q)/\partial q_n]v/f(q) \equiv g_n(q,v)$ for n = 1,...,N. We will derive these equation in more detail in section 2 below. See also section 4 in Chapter 5.

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

[11] See von Hofsten (1952), Triplett (2004), de Haan and Krsinich (2012; 31-32) and Diewert, Fox and Schreyer (2017). Inflation adjusted carry forward and backward prices were discussed in section 19 of Chapter 7.

[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.

minimizing the sum of squared errors, $\Sigma_{t=1}^{T} \Sigma_{n=1}^{N} [e_{tn}]^2 = \Sigma_{t=1}^{T} \Sigma_{n=1}^{N} [\ln p_{tn} - \ln\pi_t + \ln\alpha_n]^2$. However, the log prices are not weighted according to their economic importance and the model does not allow for missing products as was seen in Chapter 7. Finally, in section 17 of Chapter 7, we found a satisfactory stochastic model that allowed for missing observations and weighted prices by their economic importance. This model is reviewed in section 5 of the present chapter.

The model described in section 5 generates price levels that have some good axiomatic properties but the model 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 sections 6 and 7 do not suffer from this drawback.

Sections 6 and 7 deal with hedonic regression models that make use of information on the *characteristics* of the N products under consideration. The models in these two sections are more satisfactory than the weighted time product dummy model discussed in section 5 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 6 and 7 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 7. 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 two period case are addressed in section 8. 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 separate bilateral price indexes between the two periods. The two indexes, each of which hold tastes constant, are then averaged to form a final index.

Finally, in sections 9 and 10, two alternative methods for constructing reservation prices are discussed. In these methods, the underlying utility function is *not* assumed to be a linear function. In section 9, the reservation price model due to Feenstra (1994) is presented. This model assumes that the underlying preferences are CES (Constant Elasticity of Substitution).[13] The model presented in section 10 assumes that the underlying preferences are a certain flexible functional form (that is exact for the Fisher (1922) ideal quantity index). This model is due to Diewert and Feenstra (2017).

Section 11 offers some conclusions.

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

In this section, we 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.[14] 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

[13] See Arrow, Chenery, Minhas and Solow (1961) for the first use of this functional form in the economics literature. Chapter 5 considered alternative estimation methods for this functional form.
[14] 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.

are aggregated over time within each period so that the prices for each period are unit value prices. 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 it is equal to the sales of product n by the outlet (or group of outlets) for period t, while the corresponding 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.

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* or *utility function* f(q) exist. This function allows us to value alternative combinations of products; if $f(q^2) > f(q^1)$, 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 f(q) can also act as an *aggregate quantity level* for the vector of purchases, q. Thus $f(q^t)$ can be interpreted as an aggregate quantity level for the period t vector of purchases, q^t, and the ratios, $f(q^t)/f(q^1)$, t = 1,..,T, can be interpreted as *fixed base quantity indexes* covering periods 1 to T.

In the following analysis, we assume that f(q) has the following properties: (i) $f(q) > 0$ if $q >> 0_N$;[15] (ii) f(q) is nondecreasing in its components; (iii) $f(\lambda q) = \lambda f(q)$ for $q \geq 0_N$ and $\lambda \geq 0$; (iv) f(q) is a continuous concave function over the nonnegative orthant. Assumption (iii), linear homogeneity of f(q), is a somewhat restrictive assumption. However, this assumption is required to ensure that the aggregate price level, P(p,q), that corresponds to f(q) does not depend on the scale of q.[16] Property (iv) will ensure that the first order necessary conditions for the budget constrained maximization of f(q) are also sufficient.

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

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

Thus the implicit price level P(p,q), which is generated by the generic price and quantity vectors, p and q, is equal to the value of purchases, p·q, deflated by the aggregate quantity level, f(q). Note that using these definitions, the product of the aggregate price and quantity levels equals the value of purchases during the period; i.e., we have $P(p,q)f(q) = p\cdot q$.

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

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

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

[15] 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 $q \neq 0_N$,

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

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

Note that if f(q) turns out to be a *linear aggregator function*, so that $f(q^t) \equiv \alpha \cdot q^t = \Sigma_{n=1}^{N} \alpha_n q_{tn}$, 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*.[17]

In order to make further progress, it is necessary to make some additional assumptions. The two additional assumptions are: (v) f(q) is once differentiable with respect to the components of q and (vi) the observed strictly positive quantity vector for period t, $q^t >> 0_N$,[18] is a solution to the following period t constrained maximization problem: [19]

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

The first order conditions for solving (4) for period t are the following conditions:[20]

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

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

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

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:

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

Thus if we assume utility maximizing behavior on the part of purchasers of the N products using the collective utility function f(q) that satisfies the above regularity conditions, then the period t quantity aggregate is $Q^t \equiv f(q^t)$ 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

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

[17] See section 10 of Chapter 7 for the properties of quality adjusted unit value indexes.

[18] The assumption that $q^t >> 0_N$ can be replaced by the assumptions $q^t > 0_N$ and $p^t \cdot q^t > 0$.

[19] 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 chapter is the application of this theory to hedonic regression models.

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

Substitute equations (8) into equations (5) and after a bit of rearrangement, the following *fundamental equations* are obtained:[21]

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

In the following section, we will assume that the aggregator function, f(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. In the final sections of the chapter, the assumption that f(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).[22] We assume that purchasers of the three products behave as if they collectively maximized the following linear aggregator function:

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

where the α_n are positive constants. Under these assumptions, equations (10) written out in scalar form become the following equations:[23]

(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*

[21] Multiply the right hand side of equation t in (10) by $1 = Q^t/f(q^t)$ and use $P^tQ^t = v^t$ to obtain the following system of equations: $p^t = v^t\nabla_q f(q^t)/f(q^t)$ 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 are consistent with the observed period t demands q^t as functions of p^t and period t expenditure v^t. Konüs and Byushgens (1926) obtained a system of equations that is equivalent to this system of inverse demand functions. Linear homogeneity of the utility function is required in order to obtain these equations and the equivalent equations defined by (9) and (10).

[22] 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.

[23] This is a special case of the Time Product Dummy regression model which was studied in Chapter 7 and will be summarized in section 5 below. 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 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.

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*.[24]

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*;[25] 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.[26]

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.[27]

---

[24] Keynes (1930; 94) was an early author who 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*.

[25] 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.

[26] 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.

[27] 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), Diewert, Fox and Schreyer (2017) and section 19 of Chapter 7.

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 $q^1 \equiv [q_{11}, q_{12}, 0]$ 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:

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

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.[28]

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.

Define the aggregator function f(q) as follows:

$$(17)\ f_{KBF}(q^*) \equiv [q^* \cdot Aq^*]^{1/2} \equiv [\Sigma_{i=1}^N \Sigma_{j=1}^N a_{ij} q_i^* q_j^*]^{1/2}$$

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.[29] Suppose further 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 $\max_q \{f_{KBF}(q) : p^{t*} \cdot q = v^{t*} \ ; q \geq 0_N\}$ for t = 1,2 where $v^{t*} \equiv p^{t*} \cdot q^{t*}$ is the 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:

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

[28] 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.

[29] Thus $A = A^T$ and 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 $f(q^*)$ is labeled as $f_{KBF}(q^*)$. For further discussion of the regularity conditions on $f_{KBF}(q^*)$, see Diewert (1976) and Diewert and Hill (2010) or section 5 of Chapter 5.

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

Thus $Q^{2*}/Q^{1*} = f_{KBF}(q^{2*})/f_{KBF}(q^{1*})$ 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:

(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 either the Lowe, Laspeyres or Paasche index in place of the Fisher price index.[30]

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:

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

where $f_{KBF}(q^*)$ is the KBF aggregator function defined by (17) and α_1 is set equal to 1.[31] 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 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. Weighted Time Product Dummy Regressions

In this section, we consider a special case of the model of economic behavior explained in section 2 above where there are N products in the aggregate and 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 f(q) is the following *linear function*:

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

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) implies 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.

[30] 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.

[31] 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.

Note that equations (28) form the basis for the *time dummy hedonic regression model* which is due to Court (1939).[32] 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)[33] 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 approach is obviously based on only consumer preferences. This consumer oriented approach was endorsed by Griliches (1971; 14-15), Muellbauer (1974; 988) and Diewert (2003a) (2003b).[34] 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 following Court (1939), we assume that the exact model defined by (28) holds only to some degree of approximation and so we could add error terms e_{tn} to the right hand sides of equations (28). The unknown parameters, $\pi \equiv [\pi_1,...,\pi_T]$ and $\alpha \equiv [\alpha_1,...,\alpha_N]$, could be estimated as solutions to the following (nonlinear) least squares minimization problem:

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

However, in section 13 of Chapter 7, we showed that the estimated price levels π_t^* that solve the minimization problem (29) had unsatisfactory axiomatic properties. Thus we took logarithms of both sides of the exact equations (28) and added error terms to the resulting equations. This led to the following least squares minimization problem:[35]

[32] 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.

[33] "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).

[34] 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.

[35] This model is an adaptation of Summer's (1973) country product dummy model to the time series context. See Aizcorbe, Corrado and Doms (2000) for an early application of this model in the time series context.

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

where the new parameters ρ_t and β_n were defined as the logarithms of the π_t and α_n; i.e., define :

(31) $\rho_t \equiv \ln\pi_t$; t = 1,...,T;
(32) $\beta_n \equiv \ln\alpha_n$; n = 1,...,N.

However, the least squares minimization problem defined by (30) does not weight the log price terms $[\ln p_{tn} - \rho_t - \beta_n]^2$ by their *economic importance* and so in section 15 of Chapter 7, we considered the following *weighted least squares minimization problem*:[36]

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

where s_{tn} is the expenditure share of product n in period t. The first order necessary conditions for $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ to solve (33) simplify to the following T equations (34) and N equations (35):

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

The solution to (34) and (35) is not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (34) and (35), 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 (35) and drop the first equation in (34) and use linear algebra to find a unique solution for the resulting equations.[37] Once the solution is found, define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* as follows:

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

The price levels π_t^* defined by (36) are called the *Weighted Time Product Dummy price levels*. Note that the resulting *price index* between periods t and τ is equal to the following expression:

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

If $s_{tn} = s_{\tau n}$ for n = 1,...,N, then π_t^*/π_τ^* 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.

Once the estimates for the π_t and α_n have been computed, we have 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

[36] Rao (1995) (2004) (2005; 574) was the first to consider this model using expenditure share weights; see also Diewert (2004). However, Balk (1980; 70) suggested this class of models much earlier using somewhat different weights. For the case of 2 periods, see Diewert (2004) (2005a) and de Haan (2004a).
[37] 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 (34) and (35) will also generate a solution to these equations and the solution can be normalized so that $\rho_1 = 0$.

to form the aggregates using equations (38) or the α_n^* estimates can be used to form the aggregates using equations (39):[38]

$$(38)\quad P^{t*} \equiv \pi_t^* ; \quad Q^{t*} \equiv p^t \cdot q^t/\pi_t^* ; \qquad t = 1,...,T;$$
$$(39)\quad Q^{t**} \equiv \alpha^* \cdot q^t ; P^{t**} \equiv p^t \cdot q^t/\alpha^* \cdot q^t ; \qquad t = 1,...,T.$$

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 use to form period t price and quantity levels, (38) or (39).[39]

It is straightforward to generalize the weighted least squares minimization problem (33) to the case where there are missing prices and quantities. As in section 17 of Chapter 7, 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 (33) to the case of missing products is the following *weighted least squares minimization problem*:

$$(40)\quad \min_{\rho,\beta} \Sigma_{t=1}^T \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho,\beta} \Sigma_{n=1}^N \Sigma_{t\in S^*(n)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2.$$

Note that there are two equivalent ways of writing the least squares minimization problem; the first way uses the definition for the set of products n present in period t, S(t), while the second way uses the definition for the set of periods t where product n is present, $S^*(n)$. The first order necessary conditions for $\rho_1,...,\rho_T$ and $\beta_1,...,\beta_N$ to solve (40) are the following counterparts to (34) and (35):

$$(41)\quad \Sigma_{n\in S(t)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{n\in S(t)} s_{tn}\ln p_{tn} ; \qquad t = 1,...,T;$$
$$(42)\quad \Sigma_{t\in S^*(n)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{t\in S^*(n)} s_{tn}\ln p_{tn} ; \qquad n = 1,...,N.$$

As usual, the solution to (41) and (42) is not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (41) and (42), 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 (42), drop the first equation in (41) and use linear algebra to find a unique solution for the resulting equations.[40]

Define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* by definitions (31) and (32). Substitute these definitions into equations (41) and (42). After some rearrangement, equations (41) and (42) become the following ones:

[38] Note that the price level P^{t**} defined in (39) is a quality adjusted unit value index of the type studied by de Haan (2004b).
[39] In section 21 of Chapter 7, the following multilateral test was considered: *Test 2*: *The fixed basket test for prices* or *the strong identity test for quantities*: If $q^r = q^t \equiv q$, then the price index for period t relative to period r is $p_M^t(P,Q)/p_M^r(P,Q) = p^t \cdot q/p^r \cdot q$. If the price and quantity aggregates are formed using equations (39) rather than (38), then this Test will be satisfied. However, the more usual approach is to define the period t price and quantity aggregates using equations (38). If this is done, then in general, the Weighted Time Product Dummy price level functions, $p_{WTPD}^t(P,Q)$, will not satisfy the basket test, Test 2.
[40] The resulting system of T – 1 + N equations needs to be of full rank in order to obtain a unique solution.

$$(43)\ \pi_t^* = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]\ ; \qquad t = 1,...,T;$$
$$(44)\ \alpha_n^* = \exp[\Sigma_{t\in S^*(n)}\, s_{tn}\ln(p_{tn}/\pi_t^*)/\Sigma_{t\in S^*(n)}\, s_{tn}]\ ; \qquad n = 1,...,N.$$

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. The counterparts to definitions (38) are the following definitions:

$$(45)\ P^{t*} \equiv \pi_t^* = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]\ ; \qquad t = 1,...,T;$$
$$(46)\ Q^{t*} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/P^{t*}\ ; \qquad t = 1,...,T.$$

Thus P^{t*} is a weighted geometric mean of the quality adjusted prices p_{tn}/α_n^* that are present in period t where the weight for p_{tn}/α_n^* is the corresponding period t expenditure (or sales) share for product n in period t, s_{tn}. The counterparts to definitions (39) are the following definitions :

$$(47)\ Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn}\ ; \qquad t = 1,...,T;$$
$$\begin{aligned}(48)\ P^{t**} &\equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/Q^{t**} && t = 1,...,T;\\ &= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^* q_{tn} && \text{using (47)}\\ &= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*(p_{tn})^{-1}p_{tn}q_{tn}\\ &= [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_n^*)^{-1}]^{-1}\\ &\leq \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]\\ &= P^{t*}\end{aligned}$$

where the inequality follows from Schlömilch's inequality[41]; i.e., a weighted harmonic mean of the quality adjusted prices p_{tn}/α_n^* that are present in period t, P^{t**}, will always be less than or equal to the corresponding weighted geometric mean of the prices where both averages use the same share weights s_{tn} when forming the two weighted averages. The inequalities $P^{t**} \leq P^{t*}$ imply the inequalities $Q^{t**} \geq Q^{t*}$ for t = 1,...,T. This algebra is due to de Haan (2004b) (2010) and de Haan and Krsinich (2018; 763). 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 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.

If the estimated errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ that implicitly appear in the weighted least squares minimization problem turn out to equal 0, then the underlying model, $p_{tn} = \pi_t\alpha_n$ for t = 1,...,T, $n\in S(t)$, holds without error and thus provides a good approximation to reality. Moreover, under these conditions, P^{t*} will equal P^{t**} for all t. 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.

The solution to the weighted least squares regression problem defined by (40) can be used to generate imputed prices for the missing products. Thus if product n in period t is missing, define $p_{tn} \equiv \pi_t^*\alpha_n^*$. The corresponding missing quantity is defined as $q_{tn} \equiv 0$. Some statistical agencies use hedonic regression models to generate imputed prices for missing prices and then use these imputed prices in their chosen index number formula. This imputation procedure is an alternative to the inflation adjusted carry forward price procedure explained in section 3. From the viewpoint of the economic approach to index number theory, the section 3 procedure seems to be preferable since the Fisher index used in section 3 is a fully flexible functional form whereas the preferences that are exact for the Weighted Time Product Dummy model must be either linear in quantities or

[41] See Hardy, Littlewood and Pólya (1934; 26).

be Cobb Douglas. However, as indicated above, if the error terms in (40) are small, the missing product prices generated by the solution to (40) can be used with some confidence.

The axiomatic properties of the price level functions π_t^* generated by the solution to (40) were studied in section 21 of Chapter 7 and will be noted in the following section. One unsatisfactory property of the WTPD price levels π_t^* is the following one: a product that is available in only one period out of the T periods has no influence on the aggregate price levels π_t^*. This means that the price of a new product that appears in period T has no influence on 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 weighted time dummy hedonic regression models studied in this section.

6. 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, $f(q) \equiv \alpha \cdot q = \Sigma_{n=1}^N \alpha_n q_n$ 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 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.[42] 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, g(z) can be used for each quality adjustment factor; i.e., we have the following assumptions:

$$(49)\ \alpha_n \equiv g(z^n) = g(z_{n1},z_{n2},...,z_{nK})\ ; \qquad n = 1,...,N.$$

Thus each product n has its own unique mix of characteristics z^n but the *same function* g can be used to determine the relative utility to purchasers of the products.[43] 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:

$$(50)\ Q^t \equiv f(q^t) \equiv \Sigma_{n=1}^N \alpha_n q_{tn} = \Sigma_{n=1}^N g(z^n) q_{tn}\ ; \qquad t = 1,...,T.$$

Using our assumption of (exact) utility maximizing behavior with the linear utility function defined by (50), equations (10) become the following equations:

$$(51)\ p_{tn} = \pi_t g(z^n)\ ; \qquad t = 1,...,T;\ n \in S(t).$$

The assumption of approximate utility maximizing behavior is more realistic, so error terms need to be appended to equations (51). We also need to choose a functional form for the *quality adjustment function* or *hedonic valuation function* g(z). Consider the following functional form for the hedonic valuation function:

---

[42] 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.

[43] In this section, we require that each of the N products possess a positive amount of each characteristic; i.e., we require that $z^n >> 0_K$ for n = 1,...,N. This assumption will be relaxed in the following section.

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

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

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

where we have used assumptions (50) and (53). Now take logarithms of both sides of equations (51) and add error terms e_{tn} to the resulting equations. Using equations (53), we obtain the following system of estimating equations:[44]

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

where as usual, we have defined ρ_t as $\ln\pi_t$ for t = 1,...,T. Equations (54) are the equations which characterize the classic *log linear time dummy hedonic regression model*.[45] 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, which sets the error terms equal to zero, assumes that the N products are perfect substitutes once they have been quality adjusted, where the logarithms of the quality adjustment factors are defined by (53).[46]

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 (54). This leads to the following least squares minimization problem:

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

A solution ρ, γ to the minimization problem (55) will satisfy the following first order conditions:

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

Equations (56)-(58) 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 (56)-(58), 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

[44] If both sides of equation tn in equations (54) are differentiated with respect to $\ln z_{nk}$, we find that $\partial \ln p_{tn}/\partial \ln z_{nk} = \gamma_k$ 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.
[45] This model was first introduced by Court (1939) as his hedonic suggestion number 2. It was popularized by Griliches (1971; 7) and others. See Triplett (2004) and Aizcorbe (2014) for hundreds of references to the literature on the use of this model.
[46] 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 (54) means that the underlying economic model provides a closer approximation to actual behavior.

(56)-(58).[47] Choose the normalization $\rho_1^* = 0$ which is equivalent to $\pi_1^* = 1$. Thus set $\rho_1^* = 0$ in equations (56)-(58), drop the first equation in equations (56) and solve the remaining T+K equations for $\rho_2^*,..., \rho_T^*$ and $\gamma_0^*, \gamma_1^*,..., \gamma_K^*$.[48] Once these parameters have been determined, the estimated $\beta_n^* \equiv \ln\alpha_n^*$ can be defined as follows using definitions (53):

$$(59)\ \beta_n^* \equiv \ln\alpha_n^* = \ln g(z^n) = \gamma_0^* + \Sigma_{k=1}^K \gamma_k^* \ln z_{nk}\ ; \qquad n = 1,...,N.$$

Using equations (56) evaluated at ρ^* and γ^* and definitions (59), we see that $\ln\pi_t^* \equiv \rho_t^*$ is equal to the following expression:

$$(60)\ \ln\pi_t^* = [1/N(t)]\ \Sigma_{n\in S(t)} \ln(p_{tn}/\alpha_n^*)\ ; \qquad t = 1,...,T$$

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.[49] 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 (45)-(48) above.

It is useful to compare the present time dummy hedonic regression that uses characteristics information with the time dummy product regression in the previous section where the only characteristic of each product was the product itself; i.e., recall the least squares minimization problem defined by (30). It seems that this earlier model is more general than the present model. To see this, define β_n^* by definitions (59) for n = 1,...,N. Substitute these β_n^* into the objective function for the minimization problem defined by (30) in section 5. Thus these β_n^* are feasible β_n that could be inserted into (30) but they may not be optimal; i.e., in general, we can expect the time dummy product least squares minimization problem defined by (30) to deliver a *lower* sum of squared residuals than the solution to (55) delivers. Thus we might ask at this point why consider the least squares problem (55) when, in general, the least squares problem (30) will deliver a better outcome in terms of fitting the data? The problem with (30) is that there may be *no unique solution* to the least squares minimization problem (even after setting $\rho_1 = 0$) if product turnover is rapid; i.e., if there are very few matched models in the window of observations, then the regression associated with (30) 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 unique solution to (30) is the case where every product is a new one which appears in only one period.[50] In this case, there are T + N − 1 unknown ρ_t and β_n parameters (after making one normalization) and only T observed prices. Thus the use of hedonic regressions with

---

[47] We also need the modified equations (56)-(58) 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 too big relative to N, the number of products.

[48] Alternatively, set $\rho_1 = 0$ in equations (54) and run a simple linear regression to obtain estimates for the remaining parameters.

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

[50] 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 (2015) (2016) (2020), Diewert, Huang and Burnett-Issacs (2017) and Silver (2018).

characteristics information is particularly useful in situations where there is rapid product turnover and there are relatively few matched models.

The price levels π_t^* defined by (60) are not 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:

$$(61)\ \pi_t^*/\pi_\tau^* = \prod_{n\in S(t)} (p_{tn}/p_{\tau n})^{1/N(t)}.$$

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.[51]

In the previous section, 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*:[52]

$$(62)\ \min_{\rho,\gamma} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]^2$$

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 (62) will satisfy the following first order conditions:

$$(63)\ \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0\ ; \qquad t = 1,...,T;$$
$$(64)\ \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}] = 0\ ;$$
$$(65)\ \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]\ln z_{nk} = 0\ ; \qquad k = 1,...,K.$$

Equations (63)-(65) 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 (63)-(65), 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 (64)-(65).[53] Choose the normalization $\rho_1^* = 0$ which is equivalent to $\pi_1^* = 1$. Thus set $\rho_1^* = 0$ in equations (63)-(65), drop the first equation in equations (63) and solve the remaining T+K equations for ρ_2^*,..., ρ_T^* and γ_0^*, γ_1^*,..., γ_K^*. Once these parameters have been determined, the estimated β_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

---

[51] As in section 5, we note that if the estimated squared residuals for this model are small, then the estimated π_t^* defined by (60) 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.

[52] Diewert (2003b) (2005b) considered this model for the bilateral case where T = 2. Silver and Heravi (2005) and de Haan and Krsinich (2014) (2018) considered the general model.

[53] 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.

defined, the corresponding quality adjustment factors are defined as $\alpha_n^* \equiv \exp[\beta_n^*] > 0$ for n = 1,...,N.

Using equations (63) evaluated at ρ^* and γ^*, we see that $\pi_t^* \equiv \exp[\rho_t^*]$ is equal to the following expression:[54]

$$(66)\ \pi_t^* = \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]\ ; \qquad t = 1,...,T$$

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 $1 \leq t, \tau \leq T$. Note that (62) depends on the availability of expenditure share information. If, in addition, 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 (45)-(48) above.

The new price indexes are a clear improvement over their unweighted counterparts defined earlier by equations (60). In the present situation, using equations (66), 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 of each product counts in the present model whereas it did not in the corresponding unweighted model.

Note that equations (66) are the same as equations (43) in the previous section. The new quality adjustment parameters α_n^* are defined by the following counterparts to equations (44):

$$(67)\ \alpha_n^* \equiv \exp[\gamma_0^* + \Sigma_{k=1}^K\, \gamma_k^*\ln z_{nk}]\ ; \qquad n = 1,...,N.$$

Now use definitions (45)-(48) to define P^{t*}, Q^{t*}, P^{t**} and Q^{t**} where the new π_t^* and α_n^* are defined by (66) and (67). We can again deduce the inequality in (48) using these new definitions; i.e., we get the following inequalities due to de Haan (2004b) (2010) and de Haan and Krsinich (2018; 763):

$$(68)\ P^{t**} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*q_{tn} \leq \pi_t^* \equiv P^{t*}\ ; \qquad t = 1,...,T.$$

As in the previous section, P^{t*} is a weighted geometric mean of the quality adjusted prices p_{tn}/α_n^* that are present in period t where the weight for p_{tn}/α_n^* is the period t expenditure (or sales) share for product n in period t, s_{tn}, and P^{t**} is the corresponding weighted harmonic mean of the quality adjusted prices p_{tn}/α_n^* using the same weights.

The solution to the weighted least squares minimization problem defined by (62) along with the normalization $\rho_1 = 0$ can also be obtained by running the following linear regression with ρ_1 set equal to zero :

$$(69)\ (s_{tn})^{1/2}\ln p_{tn} = (s_{tn})^{1/2}\rho_t + (s_{tn})^{1/2}\gamma_0 + (s_{tn})^{1/2}\Sigma_{k=1}^K\, \gamma_k\ln z_{nk} + e_{tn}\ ; \qquad t = 1,...,T;\ n\in S(t).$$

[54] These equations are equivalent to equations (8) in de Haan and Krsinich (2018; 760).

The solution to the weighted least squares regression problem defined by (62) can be used to generate imputed prices for the missing products. Thus if product n in period t is missing, define $p_{tn} \equiv \pi_t^*\alpha_n^*$. The corresponding missing quantity is defined as $q_{tn} \equiv 0$. As was mentioned in the previous section, some statistical agencies use hedonic regression models to generate imputed prices for missing prices and then use these imputed prices in their chosen index number formula. If the weighted sum of squared errors, $\Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]^2$, is small or equivalently if the R^2 for the linear regression defined by (69) is large, then using the imputed prices generated by this model to fill in for missing prices is justified.

Using the solution functions for the price levels π_t^* given by (66) plus the definition of the weighted least squares minimization problem (62), it can be shown that π_t^* regarded as a function of $P \equiv [p^1,...,p^T]$, $Q \equiv [q^1,...,q^T]$ and $Z \equiv [z^1,...,z^K]$ satisfies the following tests:[55]

Test 1: *The weak identity test for prices*. If $p^\tau = p^t$ and $q^\tau = q^t$, then $\pi_t^*(P,Q,Z) = \pi_\tau^*(P,Q,Z)$.

Test 2: *The weak fixed basket test for prices* or *the weak identity test for quantities*. If $q^\tau = q^t \equiv q$ and $p^\tau = p^t$ then the price index for period t relative to period τ is $\pi_t^*(P,Q,Z)/\pi_\tau^*(P,Q,Z) = p^t\cdot q/p^\tau\cdot q$.

Test 3: *Linear homogeneity test for prices*. Let $\lambda > 0$. Then $\pi_t^*(p^1,...,p^{t-1},\lambda p^t,p^{t+1},...,p^T,Q,Z) = \lambda\pi_t^*(P,Q,Z)$ for $t = 1,...,T$. Thus if all prices in period t are multiplied by a common scalar factor λ, then the price level of period t relative to the price level of any other period r will increase by the multiplicative factor λ. [56]

Test 4: *Homogeneity test for quantities*. Let $\lambda > 0$. Then $\pi_t^*(P,q^1,...,q^{t-1},\lambda q^t,q^{t+1},...,q^T,Z) = \pi_t^*(P,Q,Z)$ for $t = 1,...,T$. Thus if all quantities in period t are multiplied by a common scalar factor λ, then the price level of any period r remains unchanged.

Test 5: *Invariance to changes in the units of measurement of the characteristics*. The price level functions $\pi_t^*(P,Q,Z)$ for $t = 1,...,T$ remain unchanged if the K characteristics are measured in different units.

Test 6: *Invariance to changes in the ordering of the commodities*. The price level functions $\pi_t^*(P,Q,Z)$ for $t = 1,...,T$ remain unchanged if the ordering of the N commodities is changed.

Test 7: *Invariance to changes in the ordering of the time periods*. If the T time periods are reordered by some permutation of the first T integers, then the new price level functions are equal to the same permutation of the initial price level functions.

Test 8: *Responsiveness to Isolated Products Test*: If a product is available in only one period in the window of T periods, this test asks that the price level functions $\pi_t^*(P,Q,Z)$ *respond to changes* in the prices of these isolated products; i.e., the test asks that the price level functions *are not constant* as the prices for isolated products change. This test is a variation of Test 5 suggested by Zhang, Johansen and Nygaard (2019), who suggested a bilateral version of this test.[57]

[55] See Diewert (2004) (2005b) for materials on the test approach applied to time product hedonic regressions with and without characteristics information.

[56] Furthermore, the price levels $\pi_\tau^*(P,Q,Z)$ for $\tau \neq t$ are homogeneous of degree 0 in the components of p^t; i.e., we have $\pi_\tau^*(p^1,...,p^{t-1},\lambda p^t,p^{t+1},...,p^T,Q,Z) = \pi_\tau^*(P,Q,Z)$ for all $\tau \neq t$.

[57] This test was explicitly suggested by Claude Lamboray.

The weighted time product dummy hedonic regression price levels using characteristics information, the $\pi_t^*(P,Q.Z)$, that solve (62), do *not* satisfy the following Tests 9-12.

Test 9: *The strong identity test for prices*. If $p^\tau = p^t$, then $\pi_\tau^*(P,Q,Z) = \pi_t^*(P,Q,Z)$.

Thus Test 9 is similar to Test 1 but Test 9 asks that the price levels for two periods be equal if the price vectors for the two periods are identical even if the quantity vectors for the two periods are different whereas Test 1 asks that the price levels for two periods be equal if the price *and* quantity vectors for the two periods are identical.

Test 10: *The strong fixed basket test for prices* or *the strong identity test for quantities*. If $q^\tau = q^t \equiv q$, then the price index for period t relative to period τ is $\pi_t^*(P,Q,Z)/\pi_\tau^*(P,Q,Z) = p^t \cdot q/p^\tau \cdot q$.[58]

Test 11: *Invariance to changes in the units of measurement for the quantities*. The price level functions $\pi_t^*(P,Q,Z)$ for t = 1,...,T remain unchanged if the N commodities are measured in different units of measurement.

Test 12: *Responsiveness to Changes in Imputed Prices for Missing Products Test*: If there are missing products in one or more periods, then one can define imputed prices for these missing products. This test asks that the price level functions $\pi_\tau^*(P,Q,Z)$ *respond to changes* in these imputed prices; i.e., the test asks that the price level functions *are not constant* as the imputed prices change. This test allows a price level to decline if new products enter the marketplace during the period and for consumer utility to increase as the number of available products increases. If this test is not satisfied, then the price levels will be subject to *new products bias*. This is an important source of bias in a dynamic product universe.

Many multilateral index number methods do not satisfy the strong identity Tests 9 and 10 and the responsiveness Test 12, so the failure of the hedonic regression price levels to pass these tests is not catastrophic. At first sight, the failure of the $\pi_\tau^*(P,Q,Z)$ to pass the invariance to changes in the units of measurement for the N quantities q_n is more worrisome. The failure of this test suggests that the use of hedonic regressions to adjust for quality changes *should be restricted to classes of products that are similar* and have a dominant characteristic that all of the products possess. The quantity q_n of each product should be measured in units of this dominant characteristic. Thus if the product class is candy bars, the quantity of each product should be measured by its weight. If the product class is a beverage, each product's quantity should be measured by its volume. If this advice is followed, then the unit of measurement for all quantities in the aggregate will be the same. Thus if the units of measurement change, the change of units should affect all quantities in the same way. It can be shown that the hedonic regression price levels using characteristics information, $\pi_\tau^*(P,Q,Z)$, satisfy the following test:

Test 13: *Restricted Change of Units Test*. If the units of measurement for all products are changed by the same factor, the price levels $\pi_t^*(P,Q,Z)$ remain invariant; i.e., the price levels satisfy $\pi_t^*(\delta^{-1}P,\delta Q,Z) = \pi_t^*(P,Q,Z)$ for all scalars $\delta > 0$ for t = 1,...,T.[59]

---

[58] The price levels $\pi_t^*(P,Q,Z)$ that are directly defined from the solution to (62) using equations (66) will not in general satisfy Test 10. However, if we use the solution to (62) to define the α_n^* and then use definitions (47) and (48) to define the period t price and quantity levels, P^{t**} and Q^{t**}, then the P^{t**} will satisfy Test 2. However, the present set of tests applies to the price levels $\pi_t^*(P,Q,Z)$ that are directly defined by the solution to (62).

Thus the failure of the hedonic regression price levels to pass the unrestricted change of units test, Test 6, is not catastrophic because for closely related products, these price levels will pass the restricted change of units test, Test 13.

Recall that the weighted time product dummy price levels defined in the previous section had the undesirable property that a product that is available in only 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 price levels $\pi_t^*(P,Q,Z)$ defined in this section no longer have this undesirable property since they satisfy Test 8 above.

It is possible to apply the tests listed above to the weighted time dummy price levels defined in the previous section. However, in order to do this, the g(z) function defined by (52) needs to be replaced by the linear function $g(z) \equiv \alpha \cdot z$ where z is now an N dimensional vector of characteristics (instead of a K dimensional vector). Assume that there are N models and the characteristics vector for product n is $z^n \equiv e^n$ for n = 1,...,N where e^n is the nth unit vector; i.e., e^n is an N dimensional vector which has a 1 in component n and zeros elsewhere. Thus in this case, the Z matrix is the N by N matrix $Z \equiv [z^1,z^2,...,z^N] = I_N$ where I_N is the N by N identity matrix. With this new definition for g(z) and for the matrix Z, we have $g(z^n) = g(e^n) = \alpha \cdot e^n = \alpha_n$ for n = 1,...,N, which are equations (49). Equations (51) become $p_{tn} = \pi_t g(z^n) = \pi_t \alpha_n$ for t = 1,...,T and $n \in S(t)$. From these equations, we can follow the steps in the previous section and the counterpart to the weighted least squares minimization problem (62) is (40), the final model in the previous section. Thus we can apply the above tests to the price levels that result from solving (40). We find that the weighted time dummy hedonic price levels without characteristics satisfies Tests 1-7, 11 and 13; they fail Tests 8-10 and 12. Thus the test performance of both methods is identical except that the price levels from the weighted hedonic time product dummy model that result from solving (40) pass Test 11 (invariance to changes in the units of measurement for quantities) and fail Test 8 (responsiveness to isolated products test) and the weighted hedonic time product dummy model that uses characteristics information that result from solving (62) pass Test 8 and fail Test 11.[60]

It is possible to derive some approximate equalities for the α_n^* that are counterparts to the exact equalities (44) for the α_n^* that were satisfied for the weighted time product dummy quality adjustment parameters for the model defined by (40) in the previous section. Recall that the estimated quality adjustment factors for the N products in the present model are the α_n^* defined by (67) for n = 1,...,N. The logarithms of these estimated quality adjustment factors are $\beta_n^* \equiv \ln\alpha_n^* = \gamma_0^* + \Sigma_{k=1}^K \gamma_k^* \ln z_{nk}$ for n = 1,..,N. Once the $\rho^* \equiv [\rho_1^*,\rho_2^*,...,\rho_T^*]$ and $\gamma^* \equiv [\gamma_0^*,\gamma_1^*,...,\gamma_K^*]$ solution to (62) has been determined (with $\rho_1^* = 1$), the sample residuals e_{tn}^* can be defined by equations (70) below:

$$(70)\quad \begin{aligned} e_{tn}^* &\equiv \ln p_{tn} - \rho_t^* - \gamma_0^* - \Sigma_{k=1}^K \gamma_k^* \ln z_{nk} \,; \qquad t = 1,...,T;\ n \in S(t) \\ &= \ln p_{tn} - \rho_t^* - \beta_n^* \end{aligned}$$

[59] Notation: $\delta Q = [\delta q^1,\delta q^2,..., \delta q^T]$; i.e., if the N by T matrix Q is multiplied by the scalar δ, then all NT elements in the matrix Q are multiplied by this scalar.

[60] However, as indicated earlier, often statistical agencies have to choose the hedonic regression model with characteristics over the time product dummy model explained in the previous section due to frequent model changes or to the fact that some products are unique (like housing). In the case of unique products, the time dummy approach fails and the characteristics approach is the only viable approach.

$$= \ln(p_{tn}/\pi_t^*) - \beta_n^* \qquad \text{since } \rho_t^* \equiv \ln\pi_t^*.$$

Rearranging equations (70), it can be seen that the β_n^* satisfy the following equations:

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

For each n, multiply both sides of (71) by the share s_{tn} for each $t\in S^*(n)$ and sum the resulting equations over all t that belong to the set $S^*(n)$. The following system of N equations is obtained:

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

where the approximate equalities in (72) will follow since the minimization problem defined by (62) 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^*:[61]

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

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 (73) are taken over the entire sample period.

The next few paragraphs will be devoted to addressing a problem that was first posed by de Haan and Krsinich (2018; 760): are hedonic regression models consistent with the use of unit values to aggregate over narrowly defined products at the first stage of aggregation?

Equations (70) and the definitions $\beta_n^* \equiv \ln\alpha_n^*$ for n = 1,...,N can be used to establish the following equalities:

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

Suppose that the underlying hedonic model holds exactly so that each error term e_{tn}^* is equal to 0. Finally, suppose that all of the products are *perfect substitutes* so that all of the quality adjustment factors α_n^* are equal. Thus the following equations hold:

(75) $\alpha_1^* = \alpha_2^* = ... = \alpha_N^*.$

Thus all of the estimated α_n^* will equal α_1^* for n = 2,...,N. Since the $e_{tn}^* = 0$ by assumption, $\exp[e_{tn}^*] = 1$ for t = 1,...,T; $n\in S(t)$. Substitute these relationships into equations (74). Now multiple both sides of equation tn in equations (74) 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:

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

[61] These equations provide approximate counterparts to equations (44) which were exact for the weighted time product dummy model discussed in section 5 above.

Now take ratios of equations (76) 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:

$$(77)\ \pi_t^*/\pi_1^* = \{\Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, q_{tn}\}/\{\Sigma_{n\in S(1)}\, p_{1n}q_{1n}/\Sigma_{n\in S(1)}\, q_{1n}\}; \qquad t = 1,\ldots,T.$$

The right hand side of (77) for period t can be recognized as the *unit value price index* between periods 1 and t.

The above algebra resolves the 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 (66)) 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 (74) will become $p_{tn} = \alpha_1^*\pi_t^*$ and equations (77) will hold. Thus under these conditions, there is no discontinuity problem.

As was noted above, once the estimated coefficients $\pi^* \equiv [\pi_1^*,\ldots,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_N^*]$ 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 (66) as follows:

$$(78)\ \pi_t^* = \Pi_{n=1}^N\, (p_{tn}/\alpha_n^*)^{s_{tn}}\ ; \qquad t = 1,\ldots,T.$$

Define the sequence of *hedonic price indexes*, P_H^t, as $P_H^t \equiv \pi_t^*/\pi_1^*$ for t = 1,...,T.[62] Using equations (66) and $\beta_n^* \equiv \ln\alpha_n^*$ for n = 1,...,N, we have the following expressions for the logarithms of the hedonic price indexes:

$$(79)\ \ln P_H^t = \Sigma_{n=1}^N\, s_{tn}(\ln p_{tn} - \beta_n^*) - \Sigma_{n=1}^N\, s_{1n}(\ln p_{1n} - \beta_n^*)\ ; \qquad t = 1,\ldots,T.$$

It is now possible to compare the sequence of price indexes to the corresponding Törnqvist Theil 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 Theil price index* between periods 1 and t, P_T^t, is defined as follows:[63]

$$\begin{aligned}(80)\ \ln P_T^t &\equiv \Sigma_{n=1}^N\, ½(s_{tn} + s_{1n})(\ln p_{tn} - \ln p_{1n}) \qquad t = 1,\ldots,T\\ &= \Sigma_{n=1}^N\, ½(s_{tn} + s_{1n})[(\ln p_{tn} - \beta_n^*) - (\ln p_{1n} - \beta_n^*)].\end{aligned}$$

---

[62] Recall that we set $\rho_1^* = 0$ when solving equations (63)-(65) and hence $\pi_1^* = 1$. This fact and the first equation in (66) implies that $\pi_1^* = 1 = \exp[\Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^*)] = \exp[\Sigma_{n=1}^N\, s_{1n}\ln(p_{1n}/\alpha_n^*)]$ and thus $P_H^t \equiv \pi_t^*/\pi_1^* = \pi_t^*$ 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 where π_1^* is defined by the first equation in (66).

[63] The imputed prices and shares defined above equations (78) are used to fill in any missing prices and shares in the Törnqvist formula.

Taking the difference between (79) and (80), we can derive the following expressions for t = 1,2,...,T:

(81) $\ln P_H^t - \ln P_T^t = \Sigma_{n=1}^N \frac{1}{2}(s_{tn} - s_{1n})(\ln p_{tn} - \beta_n^*) + \Sigma_{n=1}^N \frac{1}{2}(s_{tn} - s_{1n})(\ln p_{1n} - \beta_n^*)$.

Since $\Sigma_{n=1}^N (s_{tn} - s_{1n}) = 0$ for each t, the two sets of terms on the right hand side of equation t in (81) 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 $\ln p_{1n}$ 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 (81) 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_H^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.[64] 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.[65]

The hedonic valuation function g(z) defined by (49) has a useful property: one can impose constant returns to scale in the characteristics(the property $g(\lambda z) = \lambda g(z)$ 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 (63)-(65) or equations (69) as estimating equations for the unknown parameters in g(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 alternative hedonic regression model to be considered at the beginning of the following section relaxes this positivity restriction.

7. Alternative Hedonic Regression Models with Characteristics Information

As noted in the previous section, the hedonic valuation function g(z) defined by (52) 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 g(z) defined by (52) by the following functional form for g(z):

(82) $g(z_1,z_2,...,z_K) \equiv \exp[\gamma_0 + \Sigma_{k=1}^K \gamma_k z_k]$.

[64] See Diewert (2018; 39) for just such an example.

[65] 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.

Using this new hedonic valuation function and making the same assumptions (49)-(51) as were made in the previous section along with the new assumption (82), we obtain a new weighted least squares minimization problem that is a counterpart to (62). The new system of estimating equations which are counterparts to equations (69) are the following ones:

(83) $(s_{tn})^{1/2}\ln p_{tn} = (s_{tn})^{1/2}[\rho_t + \gamma_0 + \Sigma_{k=1}^{K} \gamma_k z_{nk}] + e_{tn}$; $\quad t = 1,...,T;\ n \in S(t)$

where as usual, $\rho_t \equiv \ln\pi_t$ for t = 1,...,T. We can find estimators for the unknown parameters in equations (83) by running the linear regression defined by (83) (with ρ_1 set equal to zero) or by minimizing the following sum of weighted squared residuals e_{tn} with respect to the components of the parameter vectors ρ and γ:[66]

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

A solution ρ, γ to the minimization problem (84) will satisfy the first order conditions (63)-(65) 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 (84) follows along the same lines as the share weighted model (62) defined in the previous section. In particular, in order to obtain a unique solution to the modified equations (63)-(65), we impose the normalization $\rho_1 = 0$ and drop the first equation in the modified equations (63).[67] The new product n quality adjustment parameters β_n^* and α_n^* are defined by equations (85) and the new sample residuals are defined by equations (86):

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

The new period t price levels, π_t^*, are still defined by equations (66). The remaining equations (72)-(81) in section 6 apply to the hedonic regression model defined by (84). Once the π_t^* have been calculated, the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \le t, \tau \le T$.

As usual, we can use definitions (45)-(48) to define P^{t*}, Q^{t*}, P^{t**} and Q^{t**} where the new π_t^* and α_n^* are used in these definitions. We can again deduce the de Haan inequalities $P^{t**} \le P^{t*}$ for t = 1,...,T. defined by (66) and (67). The axiomatic properties of the new price levels $\pi_t^*(P,Q,Z)$ are the same as the properties for the weighted time product dummy model that was defined by (62) in the previous section.

The hedonic regression models defined by (84) and its equally weighted counterpart which set all $s_{tn} = 1$ 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

[66] This is precisely the model studied by de Haan and Krsinich (2018). The results we derive below are identical to their results.
[67] 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.

by (84) were run for each category, the alternative price levels, P^{t*} and P^{t**}, were computed for each of the 7 categories and compared.[68] 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 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) for the last two months in the new window 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.

For our next hedonic model, we introduce a *discrete characteristic category*; i.e., each product n has a 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 continuous price determining 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 product valuation function, $g(z^n,x^n)$, that gives us the relative values for the N products where the logarithm of $g(z^n,x^n)$ is defined as follows:

[68] 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 will be explained in subsequent sections do not require information on characteristics.

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

As usual, the exact hedonic model for the prices is $p_{tn} = \pi_t g(z^n,x^n)$ 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 (87) for the N products in the sample, we obtain the following *weighted hedonic regression model*:

(88) $(s_{tn})^{1/2} \ln p_{tn} = (s_{tn})^{1/2} [\rho_t + \gamma_0 + \Sigma_{k=1}^K \gamma_k z_{nk} + \Sigma_{m=1}^M \delta_m x_{nm}] + e_{tn}$; t = 1,...,T; $n\in S(t)$.

Rather than running the above linear regression (after imposing the normalizations $\rho_1 = 0$ and $\delta_1 = 0$), we could instead minimize the following *weighted* sum of squared residuals:

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

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 (89) will satisfy the following first order conditions:

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

Equations (90)-(93) 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 (90)-(93), 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 (90)-(93), drop the first equation in equations (90), drop the first equation in (93) and solve the remaining T+K+M−1 equations for $\rho_2^*,..., \rho_T^*, \gamma_0^*, \gamma_1^*,..., \gamma_K^*, \delta_2^*,..., \delta_M^*$.[69] Once these parameters have been determined, define the estimated *logarithm of the quality adjustment factor for product n* as:

(94) $\beta_n^* \equiv \gamma_0^* + \Sigma_{k=1}^K \gamma_k^* z_{nk} + \Sigma_{m=1}^M \delta_m^* x_{nm} = \ln\alpha_n^*$; 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 (90)-(93) at the solution ρ^*, γ^*, δ^* where $\rho_1^* = 0$ and $\delta_1^* = 0$.[70] Using definitions (94), equations (90) evaluated at the above solution become the following equations:

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

---

[69] 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 (90)-(93) is assumed to be full so that the coefficient matrix has an inverse.
[70] All T+K+M+1 of the equations (90)-(93) will be satisfied at this solution.

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 the discrete product group m as $S^{**}(m)$ for m = 1,...,M. For each model n, define the *partial log adjustment factor* μ_n^* for the continuous characteristics as follows:

$$(96)\ \mu_n^* \equiv \gamma_0^* + \Sigma_{k=1}^K \gamma_k^* \ln z_{nk} ; \qquad n = 1,...,N.$$

Using these new definitions, it can be seen that equations (93), evaluated at the normalized solution to the weighted least squares minimization problem (89), can be rewritten as follows:

$$(97)\ \delta_m^* = \Sigma_{t,n\in S^{**}(m)}\, s_{tn}[\ln p_{tn} - \rho_t^* - \mu_n^*]/\Sigma_{t,n\in S^{**}(m)}\, s_{tn} ; \qquad m = 1,...,M.$$

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 (97) are intuitively plausible.

The analysis in the previous section can be adapted to the model defined by (89). Once the π_t^* have been calculated using definitions (95), the *price index* between periods t and τ is defined as π_t^*/π_τ^* for $1 \leq t, \tau \leq T$. Once the α_n^* and π_t^* have been calculated using (94) and (95), we have the usual two alternative methods for constructing period by period price and quantity levels, P^t and Q^t for t = 1,...,T. The first uses the π_t^* estimates as follows:

$$(98)\ P^{t*} \equiv \pi_t^* ; \qquad t = 1,...,T;$$
$$(99)\ Q^{t*} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/P^{t*} ; \qquad t = 1,...,T.$$

The second method uses the α_n^* estimates as follows:

$$(100)\ Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn} ; \qquad t = 1,...,T;$$
$$(101)\ P^{t**} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/Q^{t**} ; \qquad t = 1,...,T;$$

As usual, we have the inequalities $P^{t**} \leq P^{t*}$ for t = 1,...,T.

As was the case for the previous hedonic regression models, the present model can be used to generate estimates for missing prices using the equations $p_{tn} \equiv \pi_t^*\alpha_n^*$ if product n is missing in period t. Using these estimates for missing prices, the analysis below equation (81) can be used to analyse the difference between $P^{t*} = \pi_t^*/\pi_1^*$ and the Törnqvist Theil index P_T^t for period t.

We conclude this section by providing one more extension of the basic hedonic regression model using characteristics defined by (84).

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 additional amounts 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 $z_1,...,z_K$ of K continuous characteristics (where a single parameter γ_k can capture the value of an additional unit of z_k for k = 1,...,K) and the highly variable characteristic y that each product n has.

In order to obtain more flexibility with respect to the y characteristic, the observed products could be 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 parameterize this grouping, pick y^* and y^{**} such that approximately one third of the sample observations have $y \leq y^*$, one third have $y^* < y \leq y^{**}$ and one third have $y^{**} < y$. Define the following *dummy variable functions*, $D_i(y)$ for i = 1,2,3, which depend on y:

(102) $D_1(y) \equiv 1$ if $y \leq y^*$ and is equal to 0 elsewhere;
(103) $D_2(y) \equiv 1$ if $y^* < y \leq y^{**}$ and is equal to 0 elsewhere;
(104) $D_3(y) \equiv 1$ if $y^{**} < y$ and is equal to 0 elsewhere.

The above functions can be used to define the logarithm of the following *partial hedonic valuation function* h(y):

(105) $\ln h(y) \equiv D_1(y)\phi_1 y + D_2(y)[\phi_1 y^* + \phi_2(y - y^*)] + D_3(y)[\phi_1 y^* + \phi_2(y^{**} - y^*) + \phi_3(y - y^{**})]$.

Note that the logarithm of h(y) is a piecewise linear function of y.[71] If $\phi_1 = \phi_2 = \phi_3$, then $\ln h(y) = \phi_1 y$; i.e., under these conditions, lnh(y) becomes a linear function of y.

We assume the existence of an *overall hedonic valuation function*, $g(z^n,y^n)$, that defines the relative utility for the N products where product n has characteristics defined by the vector $z^n \equiv [z_{n1},...,z_{nK}]$ and the scalar y^n. The logarithm of $g(z^n,y^n)$ is defined as follows:

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

As usual, the exact hedonic model for the sample prices is $p_{tn} = \pi_t g(z^n,y^n)$ 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 (105) and (106), we obtain the following *hedonic regression model*:

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

where $\ln h(y^n)$ is defined by evaluating (105) 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 (107) in a linear fashion so the unknown parameters can be estimated using linear regression techniques.

In order to take into account the economic importance of each model, estimates for the unknown parameters in equations (107) can be obtained by minimizing the following *weighted* sum of squared residuals:

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

[71] 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. See Poirier (1976) for applications of regression models using splines.

We leave the further analysis of this model to the reader after noting that in order to obtain a unique solution to (108), 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 g(z,y) is such that lng(z,y) is linear in the unknown parameters. One can choose functions g(z,y) such that lng(z,y) 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 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.[72]

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

8. 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.[73] 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 somewhat 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, 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*[74] 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

[72] For examples of nonlinear hedonic models that make use of this nesting technique, see Chapter 10 or Diewert, Haan and Hendricks (2015), Diewert and Shimizu (2015) (2016) (2020) or Diewert, Huang and Burnett-Issacs (2017).
[73] This method is due to Court (1939) and popularized by Griliches (1971). It is called the adjacent period time dummy hedonic regression model.
[74] This rolling window time dummy hedonic model was implemented by Ivancic, Diewert and Fox (2009) and Shimizu, Nishimura and Watanabe (2010).

second window generates the sequence of price levels $\pi_2{}^2$, $\pi_3{}^2$,..., $\pi_{T+1}{}^2$. 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 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 an 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 solving the weighted least squares minimization problem defined by (84). Consider the special case of this model with only two periods so that T = 2. We reparameterize this problem defined by (84) for the case T = 2 and consider the following equivalent problem:

$$(109)\ \min_{\theta,\gamma} \Sigma_{t=1}^{2} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \theta_t - \Sigma_{k=1}^{K} \gamma_k z_{nk}]^2$$

where $\theta \equiv [\theta_1,\theta_2]$ and $\gamma \equiv [\gamma_1,...,\gamma_K]$. Comparing (109) with (84) for T = 2, it can be seen that $\theta_1 = \rho_1 + \gamma_0 = \gamma_0$ (since we set $\rho_1 = 0$ when using the model defined by (84)) and $\theta_2 = \rho_2 + \gamma_0$. Thus the two problems are completely equivalent once we impose the normalization $\rho_1 = 0$ on (84) for the case where T = 2. The first order conditions which determine a unique solution to (109)[75] are the following 2 + K equations:

$$(110)\ \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \theta_t^* - \Sigma_{k=1}^{K} \gamma_k^* z_{nk}] = 0\ ; \qquad t = 1,2;$$
$$(111)\ \Sigma_{t=1}^{2} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \theta_t^* - \Sigma_{k=1}^{K} \gamma_k^* z_{nk}]z_{nk} = 0\ ; \qquad k = 1,...,K.$$

Denote the solution to (110) and (111) by $\theta^* \equiv [\theta_1^*,\theta_2^*]$ and $\gamma^* \equiv [\gamma_1^*,...,\gamma_K^*]$. Estimates for the parameters γ_0 and ρ_2 which were used in our initial parameterization of the model defined by (84) for the case where T = 2 can be recovered from the solution to (110) and (111) as follows:[76]

[75] As usual, the coefficient matrix for the unknown parameters in equations (110) and (111) 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.
[76] The new γ_k^* are equal to the old γ_k^* for k = 1,...,K.

(112) $\gamma_0^* \equiv \theta_1^*$; $\rho_1^* \equiv 0$; $\rho_2^* \equiv \theta_2^* - \theta_1^*$.

The estimated quality adjustment parameters, β_n^* and α_n^*, for the model defined by (84) can be recovered from the estimated θ_t^* and γ_k^* by using the equations $\beta_n^* \equiv \theta_1^* + \Sigma_{k=1}^K \gamma_k^* z_{nk}$; $\alpha_n^* \equiv \exp[\beta_n^*]$ for n = 1,...,N.

However, for the remainder of this section, it will prove to be more convenient to define new quality adjustment parameters, β_n^{**} and α_n^{**}, as follows:

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

Equations (110), definitions (113) and the equations $\Sigma_{n\in S(t)} s_{tn} = 1$ for each t imply that the estimated θ_1^* and θ_2^* satisfy the following equations:

(114) $\theta_t^* = \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \Sigma_{k=1}^K \gamma_k^* z_{nk}]$ $\qquad$ t = 1,2;
$\qquad = \Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^{**})$.

Using equations (112) and (113), 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:[77]

(115) ; $\rho_2^* \equiv \theta_2^* - \theta_1^*$
$\qquad = \Sigma_{n\in S(2)} s_{2n}\ln(p_{2n}/\alpha_n^{**}) - \Sigma_{n\in S(1)} s_{1n}\ln(p_{1n}/\alpha_n^{**})$
$\qquad = \Sigma_{n\in S(2)} s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}] - \Sigma_{n\in S(1)} s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}]$.

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, $\theta_1^* + \Sigma_{k=1}^K \gamma_k^* z_{nk}$ and $\theta_2^* + \Sigma_{k=1}^K \gamma_k^* z_{nk}$, 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:

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

where the unknown parameters in (116) are θ^1, $\gamma^1 \equiv [\gamma_1^1,...,\gamma_K^1]$ and the unknown parameters in (117) are θ^2, $\gamma^2 \equiv [\gamma_1^2,...,\gamma_K^2]$. In the previous model defined by (109), there was only one vector of γ parameters to model prices in both periods while the new models defined by (116) and (117) have separate quality adjustment parameter vectors, γ^1 and γ^2.

The first order conditions for (116) are equations (118) and (119), while the first order conditions for (117) are equations (120) and (121) below:

---

[77] If the model defined by (109) held exactly so that all error terms were equal to 0, then $\ln p_{1n} = \theta_1^* + \ln\alpha_n^{**}$ for $n\in S(1)$ and $\ln p_{2n} = \theta_2^* + \ln\alpha_n^{**}$ for $n\in S(2)$. Thus $p_{1n}/\alpha_n^{**} = \exp[\theta_1^*]$ for each $n\in S(1)$ and $p_{2n}/\alpha_n^{**} = \exp[\theta_2^*]$ for each $n\in S(2)$. Thus each quality adjusted period t price, p_{tn}/α_n^{**} for $n\in S(t)$, is an estimator for $\exp[\theta_t^*]$ and thus a weighted geometric mean of these quality adjusted prices (where the weights sum to 1) is also an estimator for $\exp[\theta_t^*]$.

(118) $\Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \theta^{1*} - \Sigma_{k=1}^{K}\, \gamma_k^{1*} z_{nk}] = 0$;

(119) $\Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \theta^{1*} - \Sigma_{k=1}^{K}\, \gamma_k^{1*} z_{nk}]z_{nk} = 0$; $k = 1,...,K$;

(120) $\Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \theta^{2*} - \Sigma_{k=1}^{K}\, \gamma_k^{2*} z_{nk}] = 0$;

(121) $\Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \theta^{2*} - \Sigma_{k=1}^{K}\, \gamma_k^{2*} z_{nk}]z_{nk} = 0$; $k = 1,...,K$.

Let θ^{1*}, γ_1^{1*},..., γ_K^{1*} solve (118) and (119) and let θ^{2*}, γ_1^{2*},..., γ_K^{2*} solve (120) and (121). There are now *two sets of quality adjustment factors*: α_1^{1*},..., α_N^{1*} for period 1 and α_1^{2*},..., α_N^{2*} for period 2. The logarithms of these parameters are defined as follows:

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

Using (118), (120) and definitions (122), we obtain the following expressions for θ^{1*} and θ^{2*} as *quality adjusted log prices* for periods 1 and 2:

(123) $\theta^{1*} = \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{1*}) = \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^{K}\, \gamma_k^{1*} z_{nk}]$;

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

The average measure of log price change going from period 1 to 2 using the adjacent period time dummy hedonic model was $\rho_2^* = \theta_2^* - \theta_1^*$; see (115) 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 $\theta^{2*} - \theta^{1*}$. However, looking at (123) and (124), we see that the quality adjustment factors are not held constant in constructing this measure. The underlying exact models are now $p_{1n} = \exp[\theta^{1*}]\alpha_n^{1*}$ for $n\in S(1)$ and $p_{2n} = \exp[\theta^{2*}]\alpha_n^{2*}$ for $n\in S(2)$. 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 $\alpha_n^{1*} = \alpha_n^{2*}$. Hence 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 $p_{2n} \equiv \exp[\theta^{2*}]\alpha_n^{2*}$ and $q_{2n} = 0$. If $n\in S(2)$ but $n\notin S(1)$, define $p_{1n} \equiv \exp[\theta^{1*}]\alpha_n^{1*}$ and $q_{1n} = 0$. Using these estimated prices, we would have complete overlapping price and quantity data for the two periods. Now 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. The problem with this strategy is that the quantity index that emerges using this strategy cannot be given a welfare interpretation because preferences are allowed to change over the two periods.
- A product or model with characteristics vector $z^* \equiv [z_1^*,...,z_K^*]$ should have a log price which is approximately equal to $\theta^{1*} + \Sigma_{k=1}^{K}\, \gamma_k^{1*} z_k^* \equiv \ln p^{1*}$ in period 1 and a log price which is approximately equal to $\theta^{2*} + \Sigma_{k=1}^{K}\, \gamma_k^{2*} z_k^* \equiv \ln p^{2*}$ 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 $\ln(p^{2*}/p^{1*}) = \theta^{2*} - \theta^{1*} + \Sigma_{k=1}^{K}\, (\gamma_k^{2*} - \gamma_k^{1*})z_k^*$ can serve as a measure of average logarithmic inflation over the period. The

problem with this method is that there are many possible choices for the reference vector z^*.[78]

- 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.[79] 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 in definitions (122) 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 in definitions (122):

$$(125)\ \delta^{1*} \equiv \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{2*})\ ;\ \delta^{2*} \equiv \Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^{1*})\ .$$

It can be seen that $\theta^{2*} - \delta^{1*}$ 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, $\delta^{2*} - \theta^{1*}$ 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 (115) above.

$$\begin{aligned}(126)\ \rho_2^{**} &\equiv \tfrac{1}{2}[\theta^{2*} - \delta^{1*}] + \tfrac{1}{2}[\delta^{2*} - \theta^{1*}] \\ &= \tfrac{1}{2}[\Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^{2*}) - \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{2*})] \\ &\quad + \tfrac{1}{2}[\Sigma_{n\in S(2)}\, s_{2n}\ln(p_{2n}/\alpha_n^{1*}) - \Sigma_{n\in S(1)}\, s_{1n}\ln(p_{1n}/\alpha_n^{1*})] && \text{using (123)-(125)} \\ &= \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \tfrac{1}{2}(\ln\alpha_n^{1*} + \ln\alpha_n^{2*})] \\ &\quad - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \tfrac{1}{2}(\ln\alpha_n^{1*} + \ln\alpha_n^{2*})] \\ &= \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K (\tfrac{1}{2}\gamma_k^{1*} + \tfrac{1}{2}\gamma_k^{2*})z_{nk}] \\ &\quad - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K (\tfrac{1}{2}\gamma_k^{1*} + \tfrac{1}{2}\gamma_k^{2*})z_{nk}] && \text{using definitions (122).}\end{aligned}$$

Using (115), ρ_2^* can be expressed as follows:

$$(127)\ \rho_2^* = \Sigma_{n\in S(2)}\, s_{2n}[\ln p_{2n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}] - \Sigma_{n\in S(1)}\, s_{1n}[\ln p_{1n} - \Sigma_{k=1}^K \gamma_k^* z_{nk}].$$

The time dummy hedonic regression model defined by the minimization problem (109) 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 (116) and (117), 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 (115) or (127) and ρ_2^{**} defined by (126) is that the average hedonic coefficients $\tfrac{1}{2}\gamma_k^{1*} + \tfrac{1}{2}\gamma_k^{2*}$ are used in (126) while ρ_2^* uses the single set of coefficients γ_k^*. Thus (127) lets the single regression do the job of constructing a set of hedonic coefficients that covers both periods while (126) averages the results of the two single period regressions.

[78] Note that if γ^{1*} happens to equal γ^{2*}, then $\ln(p^{2*}/p^{1*}) = \theta^{2*} - \theta^{1*}$ and $\theta^{2*} - \theta^{1*}$ turns out to equal ρ_2^* defined by (115).

[79] The analysis which follows is due to Silver and Heravi (2007), Diewert, Heravi and Silver (2009) and de Haan (2009). For additional materials on hedonic imputation methods, see Aizcorbe (2014).

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, in general, will be the preferred approach. However, if the number of observations is small and K is relatively large, then the adjacent period time dummy approach may be less vulnerable to multicollinearity and outlier problems and hence may be the preferred approach.[80] In particular, if the number of observations for the two periods is less than 2 + 2K, then the hedonic imputation approach cannot be used. On the other hand, if the fit is very good in the two weighted least squares minimization problems defined by (115) and (116) (and there are ample degrees of freedom) and not good in the single weighted least squares minimization problem defined by (109), then it is preferable to estimate price change between the two periods using the hedonic imputation estimates for logarithmic price change defined by (126), since this difference in fit for the two models is evidence of taste change and thus it will be safer to use (126) over (127) to measure price change.

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.[81] 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.

9. 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.[82]

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:

$$(128)\quad c(p) \equiv \min{}_q\{f(q) \geq 1;\ q \geq 0_N\}.$$

[80] "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.

[81] 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.

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

If the unit cost function c(p) is known, then using duality theory, it is possible to recover the underlying utility function f(q).[83] Feenstra assumed that the unit cost function has the following *CES functional form*:

$$(129)\quad c(p) \equiv \alpha_0 \left[\sum_{n=1}^{N} \alpha_n p_n^{1-\sigma}\right]^{1/(1-\sigma)} \qquad \text{if } \sigma \neq 1;$$
$$\equiv \alpha_0 \prod_{n=1}^{N} p_n^{\alpha_n} \qquad \text{if } \sigma = 1$$

where the α_i and σ are nonnegative parameters with $\sum_{i=1}^{N} \alpha_i = 1$. The unit cost function defined by (129) is a *Constant Elasticity of Substitution (CES) utility function* which was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961)[84].

The parameter σ is the *elasticity of substitution*;[85] when $\sigma = 0$, the unit cost function defined by (129) 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 (129) 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 than the zero substitutability aggregator function (this is the special case of (129) 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 $p^t\cdot q^t \equiv \Sigma_{n=1}^{N} p_{tn}q_{tn}$. 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:

$$(130)\quad p^t\cdot q^t = u^t c(p^t)\,; \qquad t = 1,\ldots,T$$

where c is the CES unit cost function defined by (129).

[83] 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 Chapter 5 or Diewert (1974; 110-112) on the duality between linearly homogeneous aggregator functions f(q) and unit cost functions c(p).
[84] 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 Pólya (1934; 12-13). For more on estimating CES utility functions, see Chapter 5.
[85] 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 (129), $\sigma_{nk}(p) = \sigma$ 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.

Using the CES functional form defined by (129) and assuming that $\sigma \neq 1$ (or $r \neq 0$),[86] the following equations are obtained using Shephard's Lemma:

$$(131)\ q_{ti} = u^t\alpha_0 [\sum_{n=1}^N \alpha_n (p_{tn})^r]^{(1/r)-1}\alpha_i (p_{ti})^{r-1}; \qquad i = 1,\ldots,N;\ t = 1,\ldots,T$$
$$= u^t c(p^t)\, \alpha_i (p_{ti})^{r-1}/\sum_{n=1}^N \alpha_n (p_{tn})^r .$$

Premultiply equation i for period t in (131) by $p_{ti}/p^t{\cdot}q^t$. Using (129) and (131), the resulting equations can be rewritten as follows:

$$(132)\ s_{ti} = \alpha_i (p_{ti})^r/\sum_{n=1}^N \alpha_n (p_{tn})^r ; \qquad i = 1,\ldots,N;\ t = 1,\ldots,T.$$

The NT share equations defined by (132) can be used as estimating equations using a nonlinear regression approach. Note that the positive scale parameter α_0 cannot be identified using equations (132), 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 (132) are homogeneous of degree one in the parameters $\alpha_1,\ldots,\alpha_N$ 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:

$$(133)\ P_{CES}{}^t \equiv [\sum_{n=1}^N \alpha_n (p_{tn})^r]^{(1/r)} / [\sum_{n=1}^N \alpha_n (p_{1n})^r]^{(1/r)} ; \qquad t = 1,\ldots,T.$$

Suppose further that the observed price and quantity data vectors, p^t and q^t for $t = 1,\ldots,T$, satisfy equations (130) where c(p) is defined by (129) and the quantity data vectors q^t satisfy the Shephard's Lemma equations (131). 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 (129). Then Sato (1976) and Vartia (1976) showed that the sequence of CES price indexes defined by (133) *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 (132).[87] The logarithm of the period t fixed base *Sato-Vartia Index* $P_{SV}{}^t$ is defined by the following equation:

$$(134)\ \ln P_{SV}{}^t \equiv \Sigma_{n=1}^N w_n{}^t \ln(p_{tn}/p_{1n}) ; \qquad t = 1,\ldots,T.$$

The weights $w_n{}^t$ that appear in equations (134) 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 $n = 1,\ldots,N$ and $t = 1,\ldots,T$ provided that $s_{tn} \neq s_{1n}$. If $s_{tn} = s_{1n}$, then define $w_n{}^{t*} \equiv s_{tn} = s_{1n}$. The second stage weights are defined as $w_n{}^t \equiv w_n{}^{t*}/\Sigma_{i=1}^N w_i{}^{t*}$ for $n = 1,\ldots,N$ and $t = 1,\ldots,T$. Note that in order for $\ln P_{SV}{}^t$ to be well defined, we require that $s_{tn} > 0$, $s_{1n} > 0$, $p_{tn} > 0$ and $p_{1n} > 0$ for all $n = 1,\ldots,N$ and $t = 1,\ldots,T$; 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.

---

[86] When $\sigma = 1$, we have the case of Cobb-Douglas preferences. In the remainder of this section, we will assume that $\sigma > 1$ (or equivalently, that $r < 0$). This assumption means that the products under consideration are either highly substitutable (σ is considerably greater than one) or moderately substitutable (σ is greater than one but fairly close to one).

[87] See Chapter 5 for a proof of this result.

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.[88] 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.[89] 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 $(p_{tn})^r = (\infty)^r = (1/\infty)^{-r} = 0$ if r is negative.

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:

$$(135)\ c(p^t) \equiv [\sum_{n=1}^{N} \alpha_n (p_{tn})^r]^{(1/r)} = [\sum_{n\in S(t)} \alpha_n (p_{tn})^r]^{1/r} .$$

Using equations (131) for this new model with some missing products and multiplying the period t demand q_{ti} if product i is present in period t by the corresponding price p_{ti} leads to the following equations which describe the purchasers' nonzero expenditures on product i in period t:

$$\begin{aligned}(136)\ p_{ti}q_{ti} &= u^t [\sum_{n\in S(t)} \alpha_n (p_{tn})^r]^{(1/r)-1}\alpha_i (p_{ti})^r ; && t = 1,...,T;\ i\in S(t)\\ &= u^t c(p^t)\ \alpha_i (p_{ti})^r/\sum_{n\in S(t)} \alpha_n (p_{tn})^r .\end{aligned}$$

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 (135):

$$(137)\ \Sigma_{n\in S(t)}\, p_{tn}\, q_{tn} = u^t c(p^t) = u^t[\sum_{i\in S(t)} \alpha_i (p_{ti})^r]^{1/r} ; \qquad t = 1,...,T.$$

Recall that the ith sales share of product i in period t was defined as $s_{ti} \equiv p_{ti}q_{ti}/\Sigma_{n\in S(t)}\, p_{tn}q_{tn}$ for t = 1,...,T and $i\in S(t)$. Using these share definitions and equations (137), we can rewrite equations (136) in the following form:

$$\begin{aligned}(138)\ s_{ti} &= \alpha_i (p_{ti})^r/\sum_{n\in S(t)} \alpha_n (p_{tn})^r ; && t = 1,...,T;\ i\in S(t)\\ &= \alpha_i (p_{ti})^r/c(p^t)^r\end{aligned}$$

where the second set of equations follows using definitions (135).

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 (135) and

---

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

[89] 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.

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:[90]

$$(139)\ P_{CES}^t \equiv c(p^t)/c(p^1)\ ; \qquad t = 2,3,...,T$$
$$= [\sum_{i\in S(t)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)} \alpha_i (p_{1i})^r]^{1/r}$$
$$= [\text{ Index } 1]\times[\text{Index } 2]\times[\text{Index } 3]$$

where the three indexes in equations (139) are defined as follows:[91]

$$(140)\ \text{Index } 1 \equiv [\sum_{i\in S(t)\cap S(1)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_{1i})^r]^{1/r}\ ; \qquad t = 2,3,...,T;$$
$$(141)\ \text{Index } 2 \equiv [\sum_{i\in S(t)} \alpha_i (p_{ti})^r]^{1/r} / [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_{ti})^r]^{1/r}\ ; \qquad t = 2,3,...,T$$
$$(142)\ \text{Index } 3 \equiv [\sum_{i\in S(1)\cap S(t)} \alpha_i (p_{1i})^r]^{1/r} / [\sum_{i\in S(1)} \alpha_i (p_{1i})^r]^{1/r}\ ; \qquad t = 2,3,...,T.$$

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:

$$(143)\ c^{t*}(p) \equiv [\sum_{i\in S(t)\cap S(1)} \alpha_i (p_i)^r]^{1/r}\ ; \qquad t = 1,2,...,T.$$

The period t expenditure share equations defined by equations (138) using the unit cost functions defined by (143) are the following ones:

$$(144)\ s_i^{t*} \equiv p_{ti}q_{ti}/\Sigma_{n\in S(t)\cap S(1)}\, p_{tn}q_{tn} \qquad t = 1,...,T;\ i\in S(1)\cap S(t)$$
$$= \alpha_i (p_{ti})^r/\sum_{n\in S(t)\cap S(1)} \alpha_n (p_{tn})^r$$
$$= \alpha_i (p_{ti})^r/c^{t*}(p^t)^r$$

where the third equality follows using definitions (143).

Note that Index 1 is equal to $c^{t*}(p^t)/c^{t*}(p^1)$ and the Sato-Vartia formula (134) (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 $r \equiv 1-\sigma$) in order to find empirical expressions for these indexes.[92] It is convenient to define the following *observable expenditure or sales ratios*:

$$(145)\ \lambda^t \equiv \sum_{n\in S(t)} p_{tn}q_{tn}/\sum_{n\in S(1)\cap S(t)} p_{tn}q_{tn}\ ; \qquad t = 2,3,...,T;$$
$$(146)\ \mu^t \equiv \sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}\, q_{1n}\ ; \qquad t =2,3,...\ T.$$

We assume that there is at least one product that is present in periods 1 and t for each $t \geq 2$. Let product i be any one of these common products for a given $t \geq 2$. Then the share equations (138)

[90] 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 $c(p^t)/c(p^{t-1})$.
[91] The Indexes 1-3 depend on period t but we suppressed the index t from the left hand side of definitions (140)-(142).
[92] See Chapter 5 or Diewert and Feenstra (2017) for a variety of methods for estimating the elasticity of substitution.

and (144) hold for this product. These share equations can be rearranged to give us the following two sets of equations:

(147) $\alpha_i(p_{ti})^r = [\sum_{n\in S(t)} \alpha_n (p_{tn})^r]p_{ti}q_{ti}/[\sum_{n\in S(t)} p_{tn}q_{tn}]$; t = 2,3,...,T;
(148) $\alpha_i(p_{ti})^r = [\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{tn})^r]p_{ti}q_{ti}/[\sum_{n\in S(1)\cap S(t)} p_{tn}q_{tn}]$; t = 2,3,...,T.

For each t ≥ 2, equating (147) to (148) for the common product i leads to the following equations:

(149) $\sum_{n\in S(t)} \alpha_n (p_{tn})^r/\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{tn})^r = \sum_{n\in S(t)} p_{tn}q_{tn}/\sum_{n\in S(1)\cap S(t)} p_{tn}q_{tn}$; t = 2,3,...,T;
$= \lambda^t$

where the second set of equalities follows using definitions (145). Now take the 1/r root of both sides of (149) and use definitions (141) in order to obtain the following equalities:

(150) Index 2 $= [\lambda^t]^{1/r} = [\sum_{i\in S(t)} p_{ti} q_{ti}/\sum_{i\in S(1)\cap S(t)} p_{ti}q_{ti}]^{1/r}$;[93] t = 2,3,...,T.

Again assume that product i is available in periods 1 and t ≥ 2. Rearrange the share equations (138) and (144) for t = 1 and product i and we obtain the following two equations:

(151) $\alpha_i(p_{1i})^r = [\sum_{n\in S(1)} \alpha_n (p_{1n})^r]\, p_{1i}q_{1i}/[\sum_{n\in S(1)} p_{1n}q_{1n}]$; t = 2,3,...,T;
(152) $\alpha_i(p_{1i})^r = [\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{1n})^r]\, p_{1i}q_{1i}/[\sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}]$; t = 2,3,...,T.

Equating (151) to (152) leads to the following equations:

(153) $\sum_{n\in S(1)\cap S(t)} \alpha_n (p_{1n})^r/\sum_{n\in S(1)} \alpha_n (p_{1n})^r = \sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}q_{1n}$; t = 2,3,...,T;
$= \mu^t$

where the last set of equalities follows using definitions (146). Now take the 1/r root of both sides of (153) and use definitions (143) in order to obtain the following equalities:[94]

(154) Index 3 $= [\mu^t]^{1/r} = [\sum_{n\in S(1)\cap S(t)} p_{1n}q_{1n}/\sum_{n\in S(1)} p_{1n}q_{1n}]^{1/r}$; t = 2,3,...,T.

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 $[\sum_{i\in S(t)}$

---

[93] If new products become available in period t that were not available in period 1, then $\lambda^t > 1$. Recall that r = 1 − σ 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 from above is 0 and the limit of $(\lambda^t)^{1/(1-\sigma)}$ as σ approaches +∞ is 1. Thus the gains in utility from increased product variety are huge if σ is slightly greater than 1 and diminish to tiny gains as σ becomes very large. Suppose that $\lambda^t = 1.05$ 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.

[94] 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 +∞ and the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches +∞ is 1. Thus the losses in utility from decreased product variety are huge if σ is slightly greater than 1 and diminish to tiny gains as σ becomes very large. Suppose that $\mu^t = 0.95$ 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.

$p_{ti}q_{ti}/\sum_{i\in S(1)\cap S(t)} p_{ti}q_{ti}] \geq 1$. 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, $[\sum_{i\in S(1)\cap S(t)} p_{1i}q_{1i}/\sum_{i\in S(1)} p_{1i}q_{1i}] \leq 1$. If period 1 has products that are not available in period t, then the strict inequality will hold and since $1/r < 0$, 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 has *increased* in period t due to the disappearance of products that were available in period 1 but are not available in period t.

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 (132) need to be replaced by the estimating equations (138). 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) or the extensive discussion of estimation issues in Chapter 5.

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.[95] 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.[96] 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 f(q) is used in place of the CES functional form.

10. 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*, $f_{KBF}(q) \equiv [q\cdot Aq]^{1/2}$ defined by (17) in section 4.[97] 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:

$$(155)\ p^t = P^t\, \nabla_q f_{KBF}(q^t) = P^t\, [q^t\cdot Aq^t]^{-1/2}\, Aq^t\ ; \qquad t = 1,\ldots,T$$

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.

[95] 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.
[96] 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.
[97] 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.

The period t aggregate price level is P^t and the corresponding aggregate quantity level is $Q^t \equiv [q^t \cdot Aq^t]^{1/2}$ for t = 1,...,T. Multiply the right hand side of equation t in (155) by $1 = Q^t/[q^t \cdot Aq^t]^{1/2}$ for t = 1,...,T and we obtain the following system of estimating equations:

$$(156)\ p^t = P^tQ^tAq^t/q^t \cdot Aq^t = v^tAq^t/q^t \cdot Aq^t\ ; \qquad t = 1,...,T$$

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 (156). Now convert equations (156) 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:

$$(157)\ s_{tn} = \Sigma_{m=1}^{N}\ q_{tn}a_{nm}q_{tm}/\Sigma_{n=1}^{N}\ \Sigma_{m=1}^{N}\ q_{tm}a_{nm}q_{tm}\ ; \qquad t = 1,...,T;\ n = 1,...,N.$$

When estimating systems of consumer demand equations, it is common to use share equations such as equations (157) as the estimating equations. However, in our particular situation, it may be preferable to use the system of inverse demand functions defined by equations (156) as estimating equations as we shall see below.[98]

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 (156) and (157) 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 (157) 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 symmetric $A \equiv [a_{nm}]$ matrix can be determined by solving the following nonlinear least squares minimization problem:[99]

$$(158)\ \min_{A}\ \Sigma_{t=1}^{T}\ \Sigma_{n\in S(t)}\ [s_{tn} - \{\Sigma_{m=1}^{N}\ q_{tn}a_{nm}q_{tm}/\Sigma_{i=1}^{N}\ \Sigma_{j=1}^{N}\ q_{ti}a_{ij}q_{tj}\}]^2.$$

Note that the minimization problem defined by (158) is run as a single nonlinear regression rather than as a system of N share equations, which is the 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 (158), then so does λa_{nm}^* for $1 \le n \le m \le N$ for all $\lambda \ne 0$. Thus a normalization on the matrix of parameters is required for a unique solution to (158). A final point to note is that the error terms in (158) are not weighted by their economic importance. There is no need to do this because the dependent variables in (158), 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

[98] When there are missing prices, estimating systems of inverse demand functions with prices as the dependent variables is econometrically convenient. The advantages and disadvantages of alternative methods for estimating consumer preferences is discussed at some length in section 10 of Chapter 5.

[99] Alternative estimating equations are considered in Diewert and Feenstra (2017), which has a worked example. Diewert and Feenstra found that it was preferable to use the system of estimating equations (156) rather than (157) since the goal of the regressions was to find the best fitting system of inverse demand functions rather than to find the best fitting system of share equations. More research on the econometrics associated with estimating reservation prices is necessary.

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 (156) 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):

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

Note that the reservation prices defined by (159) 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:[100]

(160) $A = bb^T + B;\ b >> 0_N\ ;\ B = B^T\ ;$ 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 $f(q) = (q^Tbb^Tq)^{1/2} = (b^Tqb^Tq)^{1/2} = b^Tq$, a *linear utility function*. Thus this special case of (160) 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.[101] Write C as $[c^1,c^2,...,c^N]$

---

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

[101] $C = [c_{nk}]$ 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.

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:

$$(161)\ B = -CC^T = -\sum_{n=1}^{N} c^n c^{nT}$$

where the following restrictions on the vectors c^n are imposed in order to impose the restrictions $Bq^* = 0_N$ on B:[102]

$$(162)\ c^n \cdot q^* = 0\ ; \qquad n = 1,....,N.$$

As mentioned above, if N is not small, then usually, it will not be possible to estimate all of the parameters in the C matrix. Furthermore, frequently 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 $f(q) = b^Tq$.[103] 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 $c^1 \equiv 0_N$.[104] In the third stage, estimate $f(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2^2,...,c_N^2]$ 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.[105] 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 substitution matrix that the data will support.[106] This is the same type of procedure that Diewert and Wales (1987) (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 conceptually 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 goods problem and the problem of adjusting for quality change. However, it has the drawback of being rather complex and hence it may be resistant to large scale applications of the method. More research is required in order to develop methods that are simpler to implement.

---

[102] 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 $N-1$.

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

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

[105] 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.

[106] 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.[107]

11. Conclusion

This chapter 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 can 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[108] or the direct utility function estimation methods suggested by Diewert and Feenstra in section 10 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 statistical agencies will have to choose between these two alternative index number estimates.

[107] There is one additional method that could be used to estimate reservation prices. This method uses experimental economics to determine the price consumers would have to be paid in order to terminate their consumption of a product or service. These estimates can be converted into reservation prices; see Brynjolfsson, E., A. Collis, W.E. Diewert, F. Eggers and K.J. Fox (2019) (2020).

[108] "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 as noted above, hedonic regressions with characteristics information must be used in this situation.

- The use of weights that reflect economic importance is recommended when running hedonic regressions; see the summary of the work by de Haan and Krsinich (2018) in section 7.
- The usefulness of the weighted time product dummy hedonic regressions (without characteristics information) that was studied in section 5 is questionable; i.e., it may be preferable use the model explained in section 4 that used inflation adjusted carry forward and backward prices along with the use of a superlative index number formula for matched products.
- Weighted time dummy hedonic regression models that use characteristics information are recommended for dealing with quality adjustment problems provided that the products are moderately or highly substitutable; see sections 6 and 7.
- Section 7 developed a test approach for evaluating the properties of hedonic regressions.
- Section 8 dealt with hedonic regressions in the context of taste change. Two useful methods for estimating price levels when there is considerable product churn were suggested: adjacent period time product hedonic regressions and the hedonic imputation method. The latter method runs separate hedonic regressions for each period and averages the results of these separate regressions to obtain estimated price levels. If degrees of freedom are ample, the hedonic imputation method is recommended.
- 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 9 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 that have a downward bias.
- The econometric method explained in section 10 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 chapter 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).

Brynjolfsson, E., A. Collis, W.E. Diewert, F. Eggers and K.J. Fox (2019), "The Digital Economy, GDP and Household Welfare: Theory and Evidence", NBER Working Paper 25695, National Bureau of Economic Research, Cambridge, MA.

Brynjolfsson, E., A. Collis, W.E. Diewert, F. Eggers and K.J. Fox (2020), "Measuring the Impact of Free Goods on Real Household Consumption", Discussion Paper 20-01, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada V6T 1L4.

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

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.

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. (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. (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. (2019), "Quality Adjustment and Hedonics: A Unified Approach", Discussion Paper 19-01, 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 (2015), “Residential Property Price Indices for Tokyo”, Macroeconomic Dynamics 19, 1659-1714.

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 (2020), “Residential Property Price Indexes: Spatial Coordinates versus Neighbourhood Dummy Variables, Discussion Paper 20-01, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1.

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, 327-342.

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.

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.

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.

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.

Poirier, D.J. (1976), *The Econometrics of Structural Change*, Amsterdam: North-Holland Publishing Company.

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.

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.

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

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.

Zhang, L.-C., I. Johansen and R. Nygaard (2019), “Tests for Price Indices in a Dynamic Item Universe”, *Journal of Official Statistics*, 35:3, 683–697.