Quality Adjustment Methods

by Erwin Diewert
University of British Columbia and
University of New South Wales

17th Meeting of the Ottawa Group

Rome, Italy

June 10, 2022

Introduction

- **What is the quality adjustment problem from a statistical agency perspective?**
- **The problem boils down to this: a product is purchased in a prior period at a certain price. In the present period, a product with similar (but not exactly the same) characteristics has replaced the prior period product and it is sold at a certain price. What is the relative utility of a unit of the new product relative to a unit of the old product?**
- **It can be seen that there is no unambiguous answer to this question. There are many plausible methods for dealing with this problem.**
- **In this paper, I will attempt to place several quality adjustment methods into a common framework.**
- **This paper is Chapter 8 in the forthcoming *CPI Theory* volume. Draft chapters are available on the IMF CPI webpages.**

Introduction

- **The paper takes a consumer demand perspective to the problem of adjusting product prices for quality change.**
- **The various approaches to the problem of quality adjustment can be seen as special cases of the general framework. This framework is presented in section 2.**
- **The special cases are:**

(i) the use of inflation adjusted carry forward and carry backward prices (sections 3 and 4);

(ii)the use of hedonic regressions (sections 5-8) and

(iii) the estimation of Hicksian reservation prices (sections 9-10).

- **Due to the time constraint, I will not be able to cover the third approach in my presentation.**

The Basic Consumer Theory Framework

- **Notation: Let $p^t \equiv [p_{t1},...,p_{tN}]$ and $q^t \equiv [q_{t1},...,q_{tN}]$ denote the price and quantity vectors for time periods t = 1,...,T.**
- **The period t quantity for product n, q_{tn}, is equal to total purchases of product n by purchasers or to the sales of product n by the outlet (or group of outlets) for period t, while the period t price for product n, p_{tn}, is equal to the value of sales (or purchases) of product n in period t, v_{tn}, divided by the corresponding total quantity sold (or purchased), q_{tn}.**
- **Thus $p_{tn} \equiv v_{tn}/q_{tn}$ is the *unit value price* for product n in period t for t = 1,...,T and n = 1,...,N.**
- **Initially, we assume that all prices, quantities and values are positive; in subsequent sections, this assumption will be relaxed.**
- **I have in mind a scanner data context for an elementary category.**

The Basic Consumer Theory Framework (cont)

- **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 *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.**

Properties of $f(q^t)$

- **$f(q)$ has the following properties:**

  (i) $f(q) > 0$ if $q \gg 0_N$;

  (ii) $f(q)$ is nondecreasing in its components;

  (iii) $f(\lambda q) = \lambda f(q)$ for $q \geq 0_N$ and $\lambda \geq 0$; (linear homogeneity);

  (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, $P(p,q) \equiv p \cdot q/f(q)$ that corresponds to $f(q)$ does not depend on the scale of q.**
- **Property (iv) will ensure that the first order necessary conditions for the budget constrained maximization of $f(q)$ are also sufficient.**

The Aggregate Price Level Defined

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

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

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

More Introduction

- **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) $Q^t \equiv f(q^t)$; $t = 1,\ldots,T$.

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

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

- **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*.**

The Assumption of Maximizing Behavior is Introduced

- **Two additional assumptions are made:**

(v) f(q) is once differentiable with respect to the components of q;

(vi) the observed strictly positive quantity vector for period t,

$q^t >> 0_N$, is a solution to the following period t constrained maximization problem:

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

- **The first order conditions for solving (4) for period t are the following conditions:**

(5) $\nabla_q f(q^t) = \lambda_t p^t \; ; \quad t = 1,\ldots,T;$

(6) $p^t \cdot q^t = v^t \; ; \quad t = 1,\ldots,T.$

- **This theory dates back to Konüs and Byushgens (1926), Shephard (1953) (in the context of a cost minimization framework), Samuelson and Swamy (1974) and Diewert (1976).**

Some Implications of Maximizing Behavior

- **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) $q^t \cdot \nabla_q f(q^t) = f(q^t)$; $t = 1,\ldots,T$.

- **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) $\lambda_t = f(q^t)/p^t \cdot q^t$ $\quad\quad t = 1,\ldots,T$

$=1/P^t$

using definitions (3).

- **Thus the Lagrange multipliers for the utility maximization problems are equal to the reciprocals of the aggregate price levels.**

Additional Implications of Maximizing Behavior

- **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) $P^tQ^t = p^t \cdot q^t = v_t$; t = 1,...,T. (the product test for levels).

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

(10) $p^t = P^t \nabla_q f(q^t)$; t = 1,...,T. (Note the appearance of P^t here).

The Path Forward

- **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 (10) 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 both the hedonic regression context and in the estimation of reservation prices.**
- **Basically, different assumptions about the utility function will generate different models of quality adjustment.**
- **Since equations (10) are so important, I repeat them here!**

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

A Nonstochastic Method for Quality Adjustment: A Simple Model

- **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).**
- **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:**

The Simple 3 Product, 2 Period Model

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

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

The Simple 3 Product, 2 Period Model (cont)

- **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*.**

Reservation Prices for the Missing Prices

- **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*; i.e., they are the lowest possible prices that are just high enough to deter purchasers from purchasing the products during periods if the unavailable products hypothetically became available.**
- **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.**

Reservation Prices for the Missing Prices (cont)

- **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.**
- **The use of carry forward and backward prices to estimate missing prices is widespread in statistical agencies. For additional materials on this method for estimating missing prices, see Triplett (2004), de Haan and Krsinich (2012) and Diewert, Fox and Schreyer (2017).**
- **The simple model explained above provides a consumer theory justification for the use of these imputed prices.**

Two Methods for Computing Price and Quantity Levels

- **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, $P^1 = 1$ and P^2, can be determined.**
- **Note that $q_{13} = 0$ and $q_{22} = 0$ since consumers cannot purchase products that are not available.**
- **Use the imputed prices defined by (15) to obtain complete price vectors for each period; i.e., define the period 1 complete price vector by $p^1 \equiv [p_{11}, p_{12}, p_{13}^*]$ and the complete period 2 price vector by $p^2 \equiv [p_{21}, p_{22}^*, p_{23}]$.**
- **The corresponding complete quantity vectors are by $q^1 \equiv [q_{11}, q_{12}, 0]$ and $q^2 \equiv [q_{21}, 0, q_{23}]$.**

Two Methods for Computing Price and Quantity Levels (cont)

- 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) $Q^t \equiv \alpha \cdot q^t$; or $Q^t \equiv p^t \cdot q^t/P^t$; $\quad t = 1,2.$

- 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.

Time Product Dummy Regressions: The Case of No Missing Observations and Equal Weighting

- **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, we assume 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) $f(q) \equiv \Sigma_{n=1}^{N} \alpha_n p_n = \alpha \cdot q$

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

Time Product Dummy Regressions (cont)

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

- **where we have redefined the period t price levels P^t in equations (10) as the parameters π_t for $t = 1,\ldots,T$.**
- **Note that equations (28) form the basis for the *time dummy hedonic regression model* which is due to Court (1939). Note that these equations are a special case of the 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) and Pakes (2001) 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 consumer demands and preferences only.**

Time Product Dummy Regressions (cont)

- **Empirically, equations (28) are unlikely to hold exactly.**
- **Thus we assume that the exact model defined by (28) holds only to some degree of approximation and so error terms, e_{tn}, are added to the right hand sides of equations (28).**
- **Here are the two key questions that are addressed in the paper:**

  (i) How exactly are the error terms to be introduced into the exact equations (28)?

  (ii) Should we weight equations (28) according to their economic importance and if so, what weights should be used?
- **Our approach to answering both questions was a pragmatic one. We experimented with different ways of introducing error terms and weights into equations (28) and reject specifications which give rise to indexes which have awkward axiomatic or economic properties.**

Time Product Dummy Regressions (cont)

- **We will postpone the weighting problem for a while and look at different ways of introducing the error terms into equations (28).**
- **Our approach will not be very rigorous from an econometric point of view; we will simply generate different indexes as solutions to various least squares minimization problems.**
- **Our first approach is to simply add error terms, e_{tn}, to the right hand sides of equations (28). The unknown parameters, $\pi \equiv [\pi_1,...,\pi_T]$ and $\alpha \equiv [\alpha_1,...,\alpha_N]$, will be estimated as solutions to the following (nonlinear) least squares minimization problem:**

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

- **Throughout the paper, we obtained estimators for the aggregate price levels π_t and the quality adjustment parameters α_n as solutions to least squares minimization problems like those defined by (29) or as solutions to weighted least squares minimization problems that were considered in subsequent sections.**

Time Product Dummy Regressions: Approach 1

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

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

$= \Sigma_{t=1}^{T} \pi_t^2 (p_{tn}/\pi_t)/\Sigma_{t=1}^{T} \pi_t^2$;

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

$= \Sigma_{n=1}^{N} \alpha_n^2 (p_{tn}/\alpha_n)/\Sigma_{t=1}^{T} \alpha_n^2$.

- Solutions to the two sets of equations can readily be obtained by iterating between the two sets of equations.
- If $\pi^* \equiv [\pi_1^*,\ldots,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_N^*]$ is a solution to (30) and (31), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$.
- Then $1,\pi_2^*,\ldots,\pi_T^*$ is the sequence of price levels that is generated by the least squares minimization problem defined by (29).

Time Product Dummy Regressions: Approach 1 (cont)

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

Time Product Dummy Regressions: Approach 1 Rejected

- **From (30), it can be seen that $\alpha_n{}^*$, the quality adjustment parameter for product n, is a weighted average of the T inflation adjusted prices for product n, the $p_{tn}/\pi_t{}^*$, where the weight for $p_{tn}/\pi_t{}^*$ is $\pi_t{}^{*2} / \Sigma_{\tau=1}{}^T \pi_\tau{}^{*2}$. This means that the weight for $p_{tn}/\pi_t{}^*$ will be very high for periods t where general inflation is high, which seems rather arbitrary.**
- **In addition to having unattractive weighting properties, the estimates generated by solving the least squares minimization problem (29) suffer from a fatal flaw: *the estimates are not invariant to changes in the units of measurement*.**
- **In order to remedy this defect, we turn to an alternative error specification.**

Time Product Dummy Regressions: Approach 2

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

(29) $\ln p_{tn} = \ln\pi_t + \ln\alpha_n + \ln e_{tn}$; $\quad n = 1,...,N;\ t = 1,...,T$

$$= \rho_t + \beta_n + \varepsilon_{tn}$$

- **where $\rho_t \equiv \ln\pi_t$ for t = 1,...,T and $\beta_n \equiv \ln\alpha_n$ for n = 1,...,N.**
- **The model defined by (30*) is the basic *Time Product Dummy regression model* with no missing observations.**
- **Now choose the ρ_t and β_n to minimize the sum of squared residuals, $\Sigma_{n=1}^{N} \Sigma_{t=1}^{T} \varepsilon_{tn}^{2}$. Thus let $\rho \equiv [\rho_1,...,\rho_T]$ and $\beta \equiv [\beta_1,...,\beta_N]$ be a solution to the following least squares minimization problem:**

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

Time Product Dummy Regressions: Approach 2 (cont)

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

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

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

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

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

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

- **Thus the period t aggregate price level, π_t, is equal to the geometric average of the N quality adjusted prices for period t, $p_{t1}/\alpha_1, \ldots, p_{tN}/\alpha_N$, while the quality adjustment factor for product n, α_n, is equal to the geometric average of the T inflation adjusted prices for product n, $p_{1n}/\pi_1, \ldots, p_{Tn}/\pi_T$.**
- **These estimators look very reasonable (if quantity weights are not available).**

Time Product Dummy Regressions: Approach 2 (cont)

- If $\pi^* \equiv [\pi_1^*,\ldots,\pi_T^*]$ and $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_N^*]$ is a solution to (36) and (37), then $\lambda\pi^*$ and $\lambda^{-1}\alpha^*$ is also a solution for any $\lambda > 0$. Thus to obtain a unique solution we impose the normalization $\pi_1^* = 1$ (which corresponds to $\rho_1 = 0$).
- Then $1,\pi_2^*,\ldots,\pi_T^*$ is the sequence of price levels that is generated by the least squares minimization problem defined by (33).
- Once we have the unique solution $1,\pi_2^*,\ldots,\pi_T^*$ for the T price levels that are generated by the (33), the *price index* between period t relative to period s can be defined as π_t^*/π_s^*.
- Using equations (36) for π_t^* and π_s^*, we have the following expression for the price index:

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

- This is simply the Jevons index for period t relative to period s.

Time Product Dummy Regressions: Approach 2 (conc)

- **Thus if there are no missing observations, the Time Product Dummy price indexes between any two periods in the window of T period under consideration is equal to the *Jevons index* between the two periods (the simple geometric mean of the price ratios, p_{tn}/p_{sn}).**
- **This is a somewhat disappointing result since an equally weighted average of the price ratios is not necessarily a representative average of the prices; i.e., unimportant products to purchasers (in the sense that they spend very little on these products) are given the same weight in the Jevons measure of inflation between the two periods as is given to high expenditure products.**
- **In the following sections of the paper, we look at generalizations of the above model; i.e., we allow for missing observations and introduce weighting by economic importance.**

More General Time Product Dummy Regressions

- **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 $t = 1,2,\ldots,T$.**
- **For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t: p_{tn} > 0\}$.**
- **The generalization of (30) to the case of missing products and the use of weighting is the following *weighted least squares minimization problem:***

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

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

- **The solution to (40) is given by:**

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

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

where $\rho_t \equiv \ln\pi_t$ for $t = 1,\ldots,T$ and $\beta_n \equiv \ln\alpha_n$ for $n = 1,\ldots,N$.

- **Thus P^{t*} (the period t price level) is a share weighted average of the quality adjusted prices for period t, the p_{tn}/α_n^*.**
- **The period t quality adjusted quantity or utility level Q^{t*} is equal to period t expenditures on the product group divided by P^{t*}.**
- **On the following slide, we define Q^{t*} directly in terms of the estimated alpha parameters.**

More General Time Product Dummy Regressions (cont)

- The π_t^* estimates can be used to form the aggregates using equations (45) and (46) on the previous slide or the α_n^* estimates can be used to form the aggregates using equations (47) and (48):

(47) $Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn}$; $\quad t = 1,\ldots,T;$

(48) $P^{t**} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/Q^{t**}$; $\quad t = 1,\ldots,T;$

$= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^* q_{tn}$ $\quad$ using (47)

$= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*(p_{tn})^{-1}p_{tn}q_{tn}$

$= [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_n^*)^{-1}]^{-1}$

$\leq \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]$

$= P^{t*}$

- where the inequality follows from Schlömilch's inequality; 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).

More General Time Product Dummy Regressions (conc)

- 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.
- 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 $q_{tn} \equiv 0$.
- 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 sections 3 and 4.
- From the viewpoint of the economic approach to index number theory, the section 4 procedure seems to be preferable since the Fisher index used in section 4 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 be Cobb Douglas (in which case the expenditure shares are constant over time and there will be no missing products).
- 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.

Hedonic Regressions that Use Characteristics Information

- The new assumption in section 6 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) $\alpha_n \equiv g(z^n) = g(z_{n1},z_{n2},\ldots,z_{nK})$; $n = 1,\ldots,N$.

- 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. Define the period t quantity vector as $q^t = [q_{t1},\ldots,q_{tN}]$ for $t = 1,\ldots,T$.
- Using the above assumptions, the aggregate quantity or utility level Q^t for period t is defined as:

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

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

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

Hedonic Regressions Using Characteristics Information 2

- **Consider the following functional form for the logarithm of the $g(z^n)$:**

(52) $\ln g(z^n) = \gamma_0 + \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}$; $\quad n = 1,\ldots,N.$

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

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

where we have used assumptions (50) and (52).

- **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:**

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

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

- **Equations (54) are the equations which characterize the classic *log linear time dummy hedonic regression model*. 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.**
- **Note 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).**

Hedonic Regressions Using Characteristics Information 3

- **Estimates for $\rho \equiv [\rho_1,\ldots,\rho_T]$ and $\gamma \equiv [\gamma_0,\gamma_1,\ldots,\gamma_K]$ 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) $\min_{\rho,\gamma} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} [\ln p_{tn} - \rho_t - \gamma_0 - \Sigma_{k=1}^{K} \gamma_k \ln z_{nk}]^2.$

- **In the previous section, we noted that weighting by economic importance was preferred (but not necessary if the fit of the corresponding unweighted 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*:**

(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 $t = 1,\ldots,T$ and $n\in S(t)$. 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.**
- **Section 6 of the paper goes on to look at some of the axiomatic or test properties of the weighted hedonic regressions defined above as for the time dummy regressions defined above in section 5.**

Conclusion

The paper developed a common framework based on economic theory for dealing with quality change. Some of the conclusions from this approach are:

- **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.**
- **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., in place of this model, it may be preferable to 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.**

Bonus Slides on Clustering (1)

- **When constructing elementary indexes using scanner data, statistical agencies often find that product churn reduces the number of matched products across two time periods to a low level.**
- **In response to this lack of matching problem, one could loosen the product specification and declare that two or more products are effectively equivalent.**
- **Thus the unit value price for the elementary category that results from this grouping or clustering strategy which will be used in later stages of aggregation could suffer from unit value bias if in fact the aggregated products are not perfectly equivalent.**
- **But how could we detect possible unit value bias? One possible way is to take the price and quantity data for the products in scope and run a (weighted) time product dummy regression. If the fit is pretty good, then the prices in the group are varying in (approximately) a proportional manner and thus the prices generated by the regression will be close to any old sensible index number.**
- **If the fit is poor, then go back to the drawing board and disaggregate the products into more homogeneous groups.**
- **If the fit is poor, then another alternative is to simply give up on clustering and use a multilateral method that is base on matched models.**

Bonus Slides on Clustering (2)

- **In other words, in general, I am a fan of using a matched model methodology whenever possible (rather than using grouping or clustering of products).**
- **But I can think of some counterexamples where grouping might be called for.**

Example 1: William Nordhaus on the price of light. What counts is the number of lumens that the device generates. This allows us to compare the utility of a kerosene lamp versus an old fashion light bulb versus a new smart light bulb.

Example 2: A drug that is patented (but the patent has expired) versus a generic counterpart. The efficacy of the drug is the same for each product so a unit value aggregation seems to be called for. (Berndt and Griliches paper).

- **Both of the above examples could be treated by clustering (or by using characteristics hedonics where the main characteristic is lumens or the amount of the chemical in the drug) or by just allowing the products to be different.**
- **If we simply allow the products to be different, then we are simply back to the matched model world and we do not have to make any difficult decisions.**

Example 3: Clothing (in particular, fashion clothing). The problem here is that fashion clothing has to be treated as a seasonal product but we also need to either stratify or use hedonics so that we get some matches (versus no matches).