No. DP26-7

RCESR Discussion Paper Series

Quality Adjustment, Hedonic Regressions and the Extension Problem

March 2026

W. Erwin Diewert,
University of British Columbia and UNSW Sydney
Chihiro Shimizu, Hitotsubashi University

RCESR

The Research Center for Economic and Social Risks
Institute of Economic Research
Hitotsubashi University

2-1 Naka, Kunitachi, Tokyo, 186-8603 JAPAN
http://risk.ier.hit-u.ac.jp/

Quality Adjustment, Hedonic Regressions and the Extension Problem

W. Erwin Diewert* and Chihiro Shimizu†

March 26, 2026

Abstract

High technology products are characterized by the rapid introduction of new models and the corresponding disappearance of older models. The paper addresses the quality adjustment problem associated with the construction of price indexes for these products. A main method for dealing with this problem is the use of hedonic regression models. Hedonic regressions use either product characteristics as explanatory variables (Time Dummy Characteristics regressions) or the product itself as the ultimate characteristic (Time Product Dummy regressions). The paper considers weighted and unweighted Time Product Dummy regressions. The indexes which were generated by the hedonic regressions are compared to traditional index numbers that did not make any special adjustments for quality change. The Expanding Window variant of a Weighted Time Product Dummy regression was used to address the chain drift problem and the problems associated with extending a series that cannot be revised. Finally, the estimation of systems of inverse demand functions was also used to generate various price indexes. Seventeen alternative approaches were implemented using Japanese price and quantity data on laptop sales in Japan for the 24 months over the years 2020–2021.

Key Words: Quality adjustment, scanner data, hedonic regressions, predicted share similarity linking, expanding window approach to multilateral indexes, the chain drift problem, economic approach to index number theory, the estimation of systems of inverse demand functions.

Journal of Economic Literature Classification Codes: C32, C43, D20, D57, E31.

*Vancouver School of Economics, University of British Columbia and the School of Economics, UNSW Australia, Sydney. Email: erwin.diewert@ubc.ca.

†School of Social Data Science, Hitotsubashi University. Email: c.shimizu@r.hit-u.ac.jp.

‡The authors thank Naohito Abe, Jan de Haan, Kevin Fox, Frances Krsinich, Alicia Rambaldi, Prasada Rao and Paul Schreyer and two referees for helpful comments. The authors gratefully acknowledge financial support from the JSPS KAKEN Grant (S) 24H00012.

1 Introduction

An increasing number of business firms are willing to share their price and quantity data on their sales of consumer goods and services to a national (or international) statistical office. These data are often referred to as *scanner data*.

Some scanner data involves high technology products which are characterized by *product churn*; i.e., the rapid introduction of new models and products and the short time that these new products are sold on the marketplace. This study will look at possible methods that statistical offices could use for quality adjusting this type of data. Our empirical example will use data on the sales of laptops in Japan.

A standard method for quality adjustment is the use of hedonic regressions. These hedonic regressions regress the price of a product (or a transformation of the price) on a time dummy variable and on either a dummy variable for the product or on the amounts of the price determining characteristics of the product. The first type of model is called a *Time Product Dummy* (TPD) Hedonic regression while the second type of model is called a *Time Dummy Characteristics* Hedonic regression. We will focus on the first type of model because the most comprehensive characteristic of a product is the product itself; there is no "missing characteristic problem" with the TPD model. The theory associated with the TPD model will be discussed in section 2 below. In particular, we will relate the TPD hedonic regression model to an explicit functional form for purchaser utility functions.

Section 3 draws on the theory explained in section 2; i.e., we consider weighted and unweighted Time Product Dummy hedonic regressions in this section using data on laptops sold in Japan over a two year period. The models in this section use only a single product characteristic: the Japanese product code for each laptop sale. We consider a single panel regression versus a sequence of bilateral regressions that utilize the price and quantity data for two consecutive periods. The latter type of model can be implemented in real time and is called an *Adjacent Period Time Product Dummy* hedonic regression model.

Section 4 considers alternatives to hedonic regression models based on standard index number theory; i.e., maximum overlap chained Laspeyres, Paasche and Fisher indexes are computed in this section. We also compute the Predicted Share Similarity linked price indexes and Rolling Window GEKS indexes. The indexes calculated in this section are also "practical" indexes.

Unfortunately, the various real time indexes that are considered in sections 3 and 4 can suffer from a *chain drift problem*; i.e., an index for period t which is calculated by chaining together the results of adjacent period bilateral indexes may not be equal to the corresponding index that directly compares the prices of period 1 to the prices of period t. In section 5, drawing on the work of Chessa (2016, 2021), an *Expanding Window* variant of the Implicit Weighted Time Product Dummy index is implemented which (partially) solves the chain drift problem (but may not be suitable if the products in scope are not close substitutes). We also discuss the controversy between Krsinich (2016) and de Haan et al. (2021) on the merits of the Time Product Dummy and the Time Dummy Characteristics approaches to hedonic regressions.

Section 6 adds three more indexes that could be used to address the chain drift and lack of matching problem: the Geary Khamis index and the econometric estimation of a linear utility function and of a CES function using share equations for a system of inverse demand equations. We point out that the traditional approach to the estimation of a system of consumer demand functions that is based on estimating an expenditure function cannot be implemented in situations where there is product churn but our approach which involves the direct estimation of a utility function can work well.

Section 7 lists some tentative conclusions that we can draw from this study.

The possible contributions of the present paper to the quality change literature include the following:

- A number of possible methods for dealing with the problem of quality change in the context of constructing price indexes are compared using the same data set. In particular, hedonic regression methods are compared to the leading multilateral index number methods and to methods that involve the econometric estimation of purchaser preferences.

- In section 3, we show that unweighted indexes can be very inaccurate in situations where there is tremendous product churn.
- In section 4, we show that some multilateral methods are also subject to substantial bias in situations where there is rapid product turnover.
- In section 5, we show that the expanding window methodology pioneered by Chessa (2016, 2021) can lead to indexes which are largely free from chain drift.
- In section 6, we show how standard consumer demand theory can be adapted to deal with missing products.

2 Hedonic Regressions and Utility Theory: The Time Product Dummy Hedonic Regression Model

The theoretical framework developed in this section draws on and extends Diewert (2025), in which the first author set out the consumer-theoretic foundations of the Time Product Dummy hedonic regression model.

Quality adjustment in price index construction requires anchoring observed price movements in the utility that purchasers derive from the products in scope. Every product is characterised by a bundle of *attributes* whose presence, and amounts, shape that utility. Two broad families of *hedonic regression* have been proposed to exploit this structure. The first regresses product prices (or a suitable transformation thereof) on the measured amounts of each price-determining attribute. The second replaces the attribute vector with a set of *product dummy variables*: since every product embodies its own distinct combination of attributes, a single indicator variable encapsulates that bundle entirely.[1] Both families admit an explicit link to purchaser utility functions. Throughout this paper we work exclusively with the second class, regressing log prices on time and product indicator variables.

[1]This second class of models is the time-series counterpart of the Country-Product-Dummy (CPD) method widely used in international price comparisons; see Summers (1973) and Diewert (2005b).

Consider an economy with N products and T time periods. Write the vectors of (unit-value) prices and quantities for period t as $\mathbf{p}^t \equiv [p_{t1}, \ldots, p_{tN}]$ and $\mathbf{q}^t \equiv [q_{t1}, \ldots, q_{tN}]$, for $t = 1, \ldots, T$.[2] As a point of departure, suppose all prices and quantities are strictly positive — the case of missing products is addressed below. Our key behavioural postulate is that purchasers collectively maximise the following *linear utility function* $f(\mathbf{q})$:

$$f(\mathbf{q}) = f(q_1, q_2, \ldots, q_N) \equiv \sum_{n=1}^{N} \alpha_n q_n \equiv \boldsymbol{\alpha} \cdot \mathbf{q} \tag{2.1}$$

where each $\alpha_n > 0$ serves as a *quality adjustment factor* for product n. If all purchasers in period t face a common price vector $\mathbf{p}^t$,[3] aggregate purchases $\mathbf{q}^t$ solve $\max_{\mathbf{q}}\{\boldsymbol{\alpha} \cdot \mathbf{q} : \mathbf{p}^t \cdot \mathbf{q} = e^t;\ \mathbf{q} \geq \mathbf{0}_N\}$, with e^t denoting total expenditure in period t. The Kuhn–Tucker conditions for an interior optimum $(\mathbf{q}^t,\ \lambda_t)$ yield the following $N+1$ equations:

$$\boldsymbol{\alpha} = \lambda_t \mathbf{p}^t; \tag{2.2}$$

$$\mathbf{p}^t \cdot \mathbf{q}^t = e^t. \tag{2.3}$$

Pre-multiplying equation (2.2) by $\mathbf{q}^t$ and invoking the budget constraint (2.3) gives an explicit expression for the Lagrange multiplier:

$$\lambda_t = \boldsymbol{\alpha} \cdot \mathbf{q}^t / e^t > 0. \tag{2.4}$$

Setting $\pi_t \equiv 1/\lambda_t$ and dividing equation (2.2) through by λ_t yields the *time product dummy estimating equations*:[4]

$$p_{tn} = \pi_t \alpha_n; \quad t = 1, \ldots, T;\ n = 1, \ldots, N. \tag{2.5}$$

Period-t price and quantity aggregates are defined by $Q^t \equiv \boldsymbol{\alpha} \cdot \mathbf{q}^t$ and $P^t \equiv e^t/Q^t = \pi_t$: the price of product n in period t factorises into a common price level π_t and a product-specific quality weight α_n.

[2] This general hedonic framework follows Court (1939) and Griliches (1971).
[3] This is satisfied whenever uniform pricing prevails and quantity discounts are absent.
[4] A derivation in a related but distinct context appears in Diewert (2005b).

To handle product churn — periods in which some products are absent — let $S(t) \equiv \{n : q_{tn} > 0\}$ denote the set of products traded in period t and $S^*(n) \equiv \{t : q_{tn} > 0\}$ the set of periods in which product n appears. Log-linearising equation (2.5) and introducing expenditure-share weights s_{tn} leads to the *weighted least squares problem:*[5]

$$\min_{\boldsymbol{\rho},\boldsymbol{\beta}} \sum_{t=1}^{T} \sum_{n \in S(t)} s_{tn} \left[\ln p_{tn} - \rho_t - \beta_n\right]^2 \tag{2.6}$$

with $\rho_t \equiv \ln \pi_t$ and $\beta_n \equiv \ln \alpha_n$. Uniqueness is secured by the normalisation $\rho_1^* = 0$.

The first-order conditions of problem (2.6) admit the closed-form characterisation:

$$\pi_t^* = \exp\left[\sum_{n \in S(t)} s_{tn} \ln(p_{tn}/\alpha_n^*)\right]; \quad t = 1, \ldots, T; \tag{2.7}$$

$$\alpha_n^* = \exp\left[\frac{\sum_{t \in S^*(n)} s_{tn} \ln(p_{tn}/\pi_t^*)}{\sum_{t \in S^*(n)} s_{tn}}\right]; \quad n = 1, \ldots, N. \tag{2.8}$$

The estimated parameters (π_t^*, α_n^*) support two distinct routes to period-t price and quantity aggregates:

$$P^{t*} \equiv \pi_t^*; \quad Q^{t*} \equiv \mathbf{p}^t \cdot \mathbf{q}^t / \pi_t^*; \tag{2.9}$$

$$Q^{t**} \equiv \boldsymbol{\alpha}^* \cdot \mathbf{q}^t; \quad P^{t**} \equiv \mathbf{p}^t \cdot \mathbf{q}^t / (\boldsymbol{\alpha}^* \cdot \mathbf{q}^t) \leq P^{t*}. \tag{2.10}$$

The inequality in (2.10) is an instance of Schlömilch's inequality — a weighted harmonic mean never exceeds the corresponding weighted geometric mean.[6] In our empirical application the two routes yield virtually indistinguishable price indexes.

Not all economists subscribe to this consumer-theoretic derivation. Rosen (1974) and Triplett (1987, 2004) have advocated a more general equilibrium approach that incorporates supply-side conditions alongside demand. The purely demand-side perspective

[5] Working in logarithms is motivated by Court (1939). The unweighted (equally-weighted) counterpart yields price levels that lack invariance to the units of measurement, as shown by Diewert (2023). Economic theory naturally suggests s_{tn} as the appropriate weight.

[6] See Hardy et al. (1934, p. 26). The bounds $P^{t**} \leq P^{t*}$ were established by de Haan (2004b, 2010) and de Haan and Krsinich (2014, 2018).

adopted here has nevertheless been endorsed by Griliches (1971, pp. 14–15), Muellbauer (1974, p. 988) and Diewert (2003a,b).

3 Time Product Dummy Regression Models

We obtained data from a private firm that collects price, quantity and characteristic information on the monthly sales of laptop computers across Japan. The data are thought to cover more than 60% of all laptop sales in Japan. We utilized the data for the 24 months in the years 2021 and 2022 for our regressions and index computations. There were 2639 monthly price and quantity observations on laptops sold in total over all months. Over the 24 months in our sample, 366 distinct products were sold so $n = 1, \ldots, 366$. If each product sold in each month, we would have $366 \times 24 = 8784$ positive monthly prices and quantities, but on average, only 30.0% of the products were sold per month since $2639/8784 = 0.300$. Thus there is tremendous product churn in the sales of laptops in Japan.

The period t *average price index* P_A^t and the period t *unit value price index* P_{UV}^t are defined as follows:

$$P_A^t \equiv \frac{\sum_{n\in S(t)} p_{tn}/N(t)}{\sum_{n\in S(1)} p_{1n}/N(1)}; \quad t = 1, \ldots, 24; \tag{3.1}$$

$$P_{UV}^t \equiv \frac{\sum_{n\in S(t)} p_{tn}q_{tn}/\sum_{n\in S(t)} q_{tn}}{\sum_{n\in S(1)} p_{1n}q_{1n}/\sum_{n\in S(1)} q_{1n}}; \quad t = 1, \ldots, 24. \tag{3.2}$$

where $N(t)$ is the number of laptops sold in period t.

The Weighted Time Product Dummy least squares minimization problem was defined by (2.6). To obtain a unique solution, we added the normalization $\rho_1 = 0$. The corresponding Unweighted Time Product Dummy least squares minimization problem is defined by (2.6) with all expenditure shares s_{tn} set equal to 1. The (sample wide) unweighted Time Product Dummy regression model can be expressed as:

$$\ln \mathbf{P} = \sum_{t=2}^{24} \rho_t \mathbf{D}_t + \sum_{k=1}^{366} \beta_k \mathbf{D}_{Jk} + \mathbf{e}. \tag{3.3}$$

The R^2 for the above regression turned out to be 0.9836 where we set ρ_1^* equal to zero. The estimated ρ_t^* were exponentiated and the sequence of the $\pi_t^* \equiv \exp[\rho_t^*]$ are the *Time Product Dummy Price Indexes* P_{TPD}^t, which are listed in Table A-1 in the Appendix.

To obtain the Weighted Time Product Dummy Price Indexes, multiply the vectors on both sides of (3.3) (excluding the error vector $\mathbf{e}$) by the vector of positive square roots of the month by month expenditure shares s_{tn} on the products which were purchased in each period. The R^2 for this weighted time product dummy regression was 0.9840. The new sequence of $\pi_t^* \equiv \exp[\rho_t^*]$ are the *Weighted Time Product Dummy Price Indexes* P_{WTPD}^t, which are listed in Table A-1 and plotted on Chart 1 below. Define the *Implicit Weighted Time Product Dummy* period t price index as $P_{IWTPD}^t \equiv P^{t**}/P^{1**}$ using equations (2.10). The indexes P_{IWTPD}^t are also listed in Table A-1 and plotted on Chart 1 below.

Since P_{WTPD}^t requires data from all T periods simultaneously, it cannot be computed in real time. To obtain real-time indexes, we turn to the adjacent period methodology. Define $S(1,2)$ as the set of products that were purchased in months 1 and 2. The counterpart regression to the unweighted time product dummy hedonic regression that links the prices of months 1 and 2 is:

$$\ln \mathbf{P}^* = \rho_2^* \mathbf{D}_2^* + \sum_{k \in S(1,2)} \beta_k \mathbf{D}_{Jk}^* + \mathbf{e}^* \tag{3.4}$$

where $\ln \mathbf{P}^*$, $\mathbf{D}_2^*$, and the product dummy vectors are all restricted to the observations corresponding to the products n that were sold in periods 1 and 2.[7] Using the results of the latter regression, we defined $P_{APTPD}^1 = 1$ and $P_{APTPD}^2 = \exp[\rho_2^*]$. In all, 23 unweighted bilateral time product dummy variable regressions were run. The Adjacent Period Time Product Dummy Price Indexes P_{APTPD}^t and their weighted counterparts P_{WAPTPD}^t[8] are listed in Table A-1 in the Appendix, together with the simple average and unit value price indexes, P_A^t and P_{UV}^t. Chart 1 below plots these indexes.

[7] Due to rapid product turnover, most of the product dummy variable vectors $\mathbf{D}_{Jk}^*$ will be vectors of zeros if we include all 366 products. The second line in (3.4) sums only over the products that actually sold in periods 1 and 2.

[8] In order to obtain the Weighted Adjacent Period Time Product Dummy Price Indexes P_{WAPTPD}^t, we took the 23 bilateral regressions and multiplied the dependent and independent variables in each of these regressions by the square root of the appropriate expenditure share.

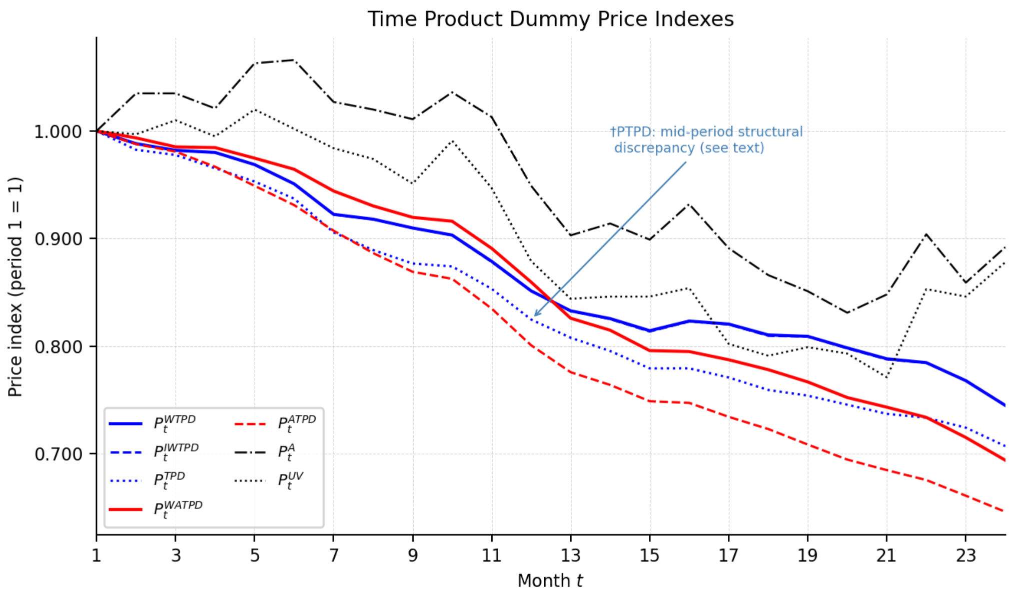

Figure 1: Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Indexes

The average price index P_A^t, and the Unit Value price index P_{UV}^t, are much higher than the other indexes because they do not adjust for improvements in the quality of laptops over time. They are also more variable than the other indexes.

There are large differences between the weighted and unweighted Time Product Dummy price indexes with the unweighted indexes generating lower rates of laptop inflation. We prefer the weighted estimates over their unweighted counterparts due to the unrepresentative nature of the unweighted indexes. However, the adjacent period indexes, P_{APTPD}^t and P_{WAPTPD}^t end up well below their panel data counterpart indexes, P_{TPD}^t and P_{WTPD}^t. Since there are large month to month fluctuations in laptop prices and quantities, it seems likely that the adjacent period indexes suffer from a chain drift problem.

4 Traditional Indexes and Similarity Linked Price Indexes for Laptops

Some of the indexes defined in the previous sections were *chained* indexes; i.e., the index constructed for month t compared the prices for month t with the prices for month $t-1$.

However, it is not the case that all bilateral comparisons of prices between two months are equally accurate: if the relative prices for matched products in months r and t are very similar, then the Laspeyres and Paasche price indexes will be close to each other and hence it is likely that the "true" price comparison between these two periods will be close to the bilateral Fisher (1922) index. This test suggests that a more accurate set of price indexes could be constructed if a bilateral comparison of prices was made between the two months that have the most *similar relative price* structures.[9] The *Predicted Share* method of linking months with the most similar structure of relative prices will be explained under the assumption that it is necessary to construct a price index P^t in real time.[10]

The price and quantity data are organized into 24 price and quantity vectors of dimension 366, $\mathbf{p}^t \equiv [p_1^t, p_2^t, \ldots, p_{366}^t]$ and $\mathbf{q}^t \equiv [q_1^t, q_2^t, \ldots, q_{366}^t]$, for $t = 1, \ldots, 24$. If product k is not purchased during month t, then we set $p_k^t = q_k^t = 0$. For months r and t, define the set of products k that are present in both months as $S(r,t)$. The *matched model Laspeyres and Paasche indexes*, $P_L(t/r)$ and $P_P(t/r)$, and the *matched model Fisher index* $P_F(t/r)$ are defined as follows:

$$P_L(t/r) \equiv \frac{\sum_{k \in S(r,t)} p_k^t q_k^r}{\sum_{k \in S(r,t)} p_k^r q_k^r}; \quad 1 \le r, t \le 24; \tag{4.1}$$

$$P_P(t/r) \equiv \frac{\sum_{k \in S(r,t)} p_k^t q_k^t}{\sum_{k \in S(r,t)} p_k^r q_k^t}; \quad 1 \le r, t \le 24; \tag{4.2}$$

$$P_F(t/r) \equiv \left[P_L(t/r) P_P(t/r)\right]^{1/2}; \quad 1 \le r, t \le 24.^{11} \tag{4.3}$$

The month t expenditure shares are:

$$s_k^t \equiv p_k^t q_k^t / (\mathbf{p}^t \cdot \mathbf{q}^t); \quad t = 1, \ldots, 24; \; k = 1, \ldots, 366. \tag{4.4}$$

[9] This approach was used by Hill (1997, 2004) and by Shankar and Hajargasht (2017) in the spatial context. Hill (2001, 2004) also pursued this similarity of relative prices approach in the time series context.

[10] This method is explained more fully in Diewert (2023).

The following *Predicted Share measure of relative price dissimilarity*, $\Delta(\mathbf{p}^r, \mathbf{p}^t, \mathbf{q}^r, \mathbf{q}^t)$, is well defined even if some product prices in the two periods being compared are equal to zero:[12]

$$\Delta(\mathbf{p}^r, \mathbf{p}^t, \mathbf{q}^r, \mathbf{q}^t) \equiv \sum_{k=1}^{366} \left[s_k^t - \frac{p_k^r q_k^t}{\mathbf{p}^r \cdot \mathbf{q}^t} \right]^2 + \sum_{k=1}^{366} \left[s_k^r - \frac{p_k^t q_k^r}{\mathbf{p}^t \cdot \mathbf{q}^r} \right]^2 ; \quad 1 \leq r, t \leq 24. \quad (4.5)$$

The 24×24 matrix of Predicted Share measures of relative price similarity for our laptop data are listed in Table A-5 in the Appendix. This matrix can be used to construct the relative price similarity linked Predicted Share Price index, P_S^t, for $t = 1, \ldots, 24$. See Diewert and Shimizu (2026) for the details of the construction.

It turns out that the relative price similarity linked indexes P_S^t, the Fisher chained maximum overlap indexes P_{FCH}^t and the Adjacent Period Weighted Time Product Dummy price indexes P_{WAPTPD}^t are all extremely close to each other for our laptop data set.[13]

There are two additional indexes that we compute: the GEKS indexes P_{GEKS}^t and the Rolling Window GEKS with a mean splice indexes P_{RWGEKS}^t.[14] The preliminary GEKS price level for period t, P_{GEKSP}^t, is defined as follows:

$$P_{GEKS}^t \equiv P_{GEKSP}^t / P_{GEKSP}^1; \quad P_{GEKSP}^t \equiv \left[\prod_{r=1}^{24} P_F(t/r) \right]^{1/24} ; \quad t = 1, \ldots, 24. \quad (4.6)$$

A problem with the GEKS indexes in our present context is the rapid introduction of new products and a corresponding rapid pace of product disappearance. Only one product (#162) was purchased in all 24 months. Thus the Fisher bilateral price index that compares the prices in month 24 to the same product prices in month 1 collapses down to the price of product 162 in month 24 divided by the price of product 162 in month 1. The GEKS index is likely to suffer from a *lack of matching bias*.

[12] See Diewert (2023) for the axiomatic properties of this measure.

[13] The bilateral link Fisher indexes that were used to construct the similarity linked indexes were equal to adjacent period matched model bilateral Fisher indexes for 20 out of 23 bilateral links. This explains why the chained Fisher index P_{FCH}^t is so close to the relative price similarity linked indexes P_S^t. The bilateral link indexes used to construct the Weighted Adjacent Period Time Product Dummy indexes P_{WAPTPD}^t are also numerically close to the corresponding matched model bilateral Fisher index, which explains why P_{WAPTPD}^t is close to P_{FCH}^t; see Diewert (2005a) on this approximation.

[14] The GEKS index dates back to Gini (1931) and the Rolling Window with a mean splice dates back to Ivancic et al. (2011) and Diewert and Fox (2022).

Table A-2 in the Appendix lists the indexes P_S^t, P_{FCH}^t, P_{LCH}^t, P_{PCH}^t, P_{GEKS}^t and P_{RWGEKS}^t, and they are plotted in Chart 2 below.

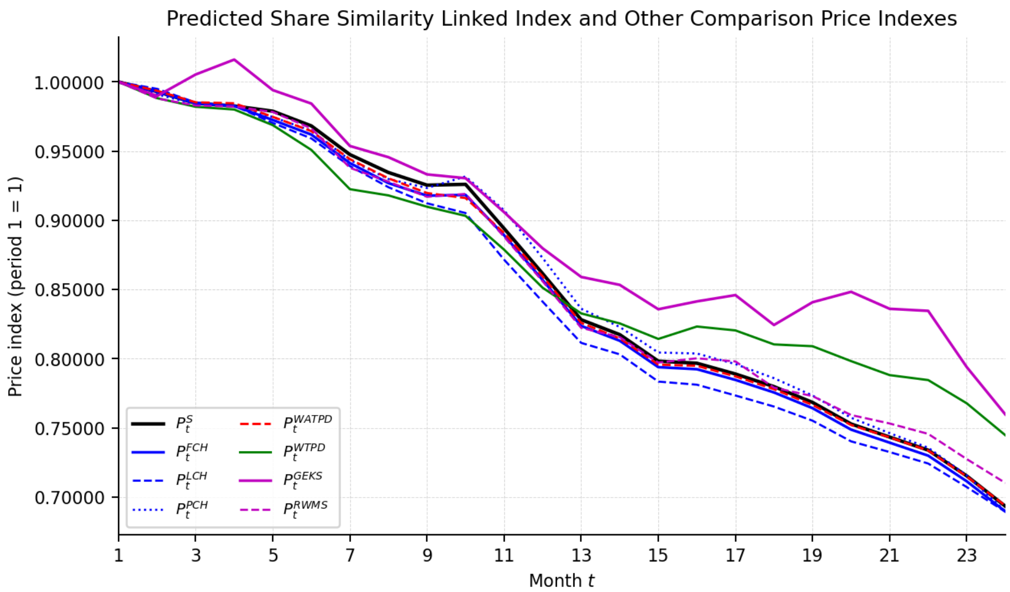

Figure 2: The Predicted Share Similarity Linked Index and Other Comparison Price Indexes

The sample wide GEKS index, P_{GEKS}^t, finishes higher than the other indexes followed by the sample wide Weighted Time Product Dummy index P_{WTPD}^t, which is virtually identical to its counterpart Implicit Time Product Dummy index, P_{IWTPD}^t (not shown). These three indexes are completely transitive and hence are not subject to chain drift. However, the sample wide GEKS indexes are not very reliable due to an extreme lack of product matching over time. It is likely that the 5 indexes P_S^t, P_{FCH}^t, P_{LCH}^t, P_{PCH}^t and P_{WAPTPD}^t all suffer from some downward chain drift. The fact that the similarity linked indexes P_S^t satisfy Walsh's (1921) Multiperiod Identity Test is not sufficient to rule out chain drift.[15]

In the following section, we suggest a new approach to the extension problem.

[15]In order to be certain to eliminate chain drift, we need the indexes to satisfy the transitivity or circularity test in index number theory; see Chapter 8 in Fisher (1922). The problem with the predicted share methodology is that it tends to pick the previous period as the preferred period to link the current period index to an earlier most similar in structure period. Fox et al. (2025) noted this problem in their study of the chain drift problem for UK prices.

5 Expanding Window Weighted Time Product Dummy Indexes

We can determine whether a given price index suffers from a chain drift problem by comparing it to a "reasonable" index that does not suffer from chain drift. Consider the sample wide Weighted Time Product Dummy price index, P^t_{WTPD}. Using equations (2.10), define month t aggregate quantity levels Q^{t**} and the Implicit Weighted Time Product Dummy price levels P^{t**}:

$$Q^{t**} \equiv \boldsymbol{\alpha}^* \cdot \mathbf{q}^t; \qquad P^{t**} \equiv \mathbf{p}^t \cdot \mathbf{q}^t/(\boldsymbol{\alpha}^* \cdot \mathbf{q}^t); \quad t = 1, \ldots, 24. \tag{5.1}$$

Define the month t *Implicit Weighted Time Product Dummy price index* P^t_{IWTPD} as:

$$P^t_{IWTPD} \equiv P^{t**}/P^{1**}; \quad t = 1, \ldots, 24. \tag{5.2}$$

The implicit price index P^t_{IWTPD} is also *transitive*; i.e., it satisfies the *circularity test* and thus it is free from chain drift.[16]

Our first suggested approximation constructs *Expanding Window Weighted Time Product Dummy price indexes*, P^t_{EW}, for $t = 1, 2, \ldots, 24$. We show how this can be done.

Step 1: Define $P^1_{EW} \equiv 1$.

Step 2: Run the weighted Time Product Dummy regression that links months 1 and 2 as in section 3 above. Exponentiate the estimated β^*_k to get estimated α^*_k for $k = 1, \ldots, 366$. Use definitions (5.1) for $t = 1, 2$ to define P^{1**} and P^{2**}. Define $P^2_{EW} \equiv P^{2**}/P^{1**}$.

Step 3: Run a weighted Time Product Dummy regression using the data for months 1, 2 and 3.[17] Define $P^3_{EW} \equiv P^{3**}/P^{1**}$.

[16] Using an expanding window (instead of using a rolling window) to construct multilateral indexes was suggested (and implemented) by Chessa (2016, 2021). The idea of using an ever expanding window was suggested by Diewert (2023) in the context of the predicted share price similarity methodology.

[17] The unweighted counterpart to the weighted regression has the form $\ln \mathbf{P}^* = \rho^*_2 \mathbf{D}^*_2 + \rho^*_3 \mathbf{D}^*_3 + \sum_{k \in S(1,2,3)} \beta_k \mathbf{D}^*_{Jk} + \mathbf{e}^*$ where $S(1, 2, 3)$ is the set of products which were sold in months 1, 2 or 3.

⋮

Step 24: The final step simply sets P_{EW}^{24} equal to the month 24 Implicit Weighted Time Product Dummy price index P_{IWTPD}^{24}.

The *Expanding Window price indexes*, P_{EW}^{t}, are listed in Table A-3 in the Appendix and plotted on Chart 3. It can be seen that the Expanding Window price indexes P_{EW}^{t} are reasonably close to their transitive counterpart indexes, the Implicit Weighted Time Product Dummy indexes, P_{IWTPD}^{t}, and they are very close near the end of the sample period.

If the statistical agency is able to collect price and quantity data on the products in scope on a retrospective basis, then Modified Expanding Window price indexes P_{MEW}^{t} could be used.[18] To construct these indexes, start off with a window of 12 months of data and construct Implicit Weighted Time Product Dummy indexes for this 12 month window. Then simply switch over to the Expanding Window price indexes for months 13 to 24. The *Modified Expanding Window price indexes*, P_{MEW}^{t}, are also listed in Table A-3 and plotted on Chart 3.

The Modified Expanding Window indexes P_{MEW}^{t} are in general closer to our target transitive indexes P_{IWTPD}^{t} for months 1–12 than the Expanding Window indexes P_{EW}^{t}. Of course, P_{MEW}^{t} coincides with P_{EW}^{t} for months 13–24 by construction. The first 3 indexes listed in Table A-3 all coincide at month 24.

The limitations of the Expanding Window Weighted TPD model include:[19]

- New and existing products must compete in the marketplace for more than one period. It is this competition between products that will enable us to estimate the relative values of the products to purchasers.[20]

[18] Krsinich (2016, p. 401) noted that estimates for the quality adjustment parameters α_n are not reliably determined until the products have been present in the marketplace for several periods. This observation helps to explain why the Modified Expanding Window (EW) estimates will be more accurate than the simple EW estimates for the first few periods.

[19] For an early application of the Rolling Window Weighted TPD model, see section 7 of Ivancic et al. (2009).

[20] This limitation of the TPD methodology was recognized by Krsinich (2016, pp. 400–401) and de Haan et al. (2021, p. 395). Consider the extreme example where a new product enters the marketplace every period but exits after only one month. There is an extreme *lack of matching bias*.

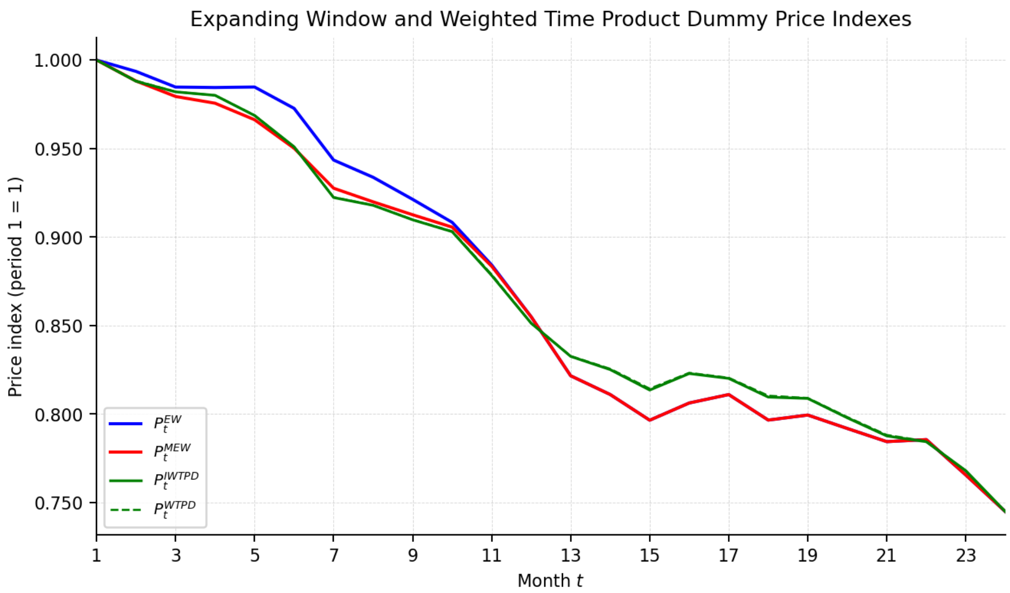

Figure 3: Expanding Window and Weighted Time Product Dummy Price Indexes

- The Time Product Dummy model relies on the assumption that purchasers have the same linear preferences over the products in scope. If the R^2 is low, it may be best to turn to a model that allows for more flexible purchaser preferences such as Rolling Year GEKS or CCDI.[21]

- The Expanding Window Weighted TPD model does not allow for preference changes.

6 Other Preference Models

In this section, we will consider only the estimation of sample wide indexes. The Ever Expanding Window methodology considered in the previous section can be adapted to produce real time Nonrevisable Indexes.

6.1 The Geary-Khamis Index

Another multilateral index that is being used by some National Statistical Offices is the Geary (1958) Khamis (1970) price index. The GK index is used in the Netherlands to construct some components of the Dutch CPI; see Chessa (2016).

[21]On the Rolling Year GEKS methodology, see Ivancic et al. (2011). On the Rolling Year CCDI methodology, see Diewert and Fox (2022) and Fox et al. (2025). The CCDI multilateral price index is due to Caves et al. (1982) and Inklaar and Diewert (2016).

Define $S^*(n)$ as the set of periods t where product n was sold. Define the vector $\mathbf{q} \equiv \sum_{t=1}^{24} \mathbf{q}^t$. The equations which determine the *GK price levels* $P_{GK}^1, \ldots, P_{GK}^{24}$ and the *quality adjustment factors* $\alpha_1, \ldots, \alpha_{366}$ (up to a scalar multiple) are:

$$\alpha_n = \sum_{t \in S^*(n)} \left[\frac{q_{tn}}{q_n}\right] \left[\frac{p_{tn}}{P_{GK}^t}\right]; \quad n = 1, \ldots, 366; \tag{6.1}$$

$$P_{GK}^t = \sum_{n \in S(t)} p_{tn} q_{tn} / (\boldsymbol{\alpha} \cdot \mathbf{q}^t); \quad t = 1, \ldots, 24 \tag{6.2}$$

where $\boldsymbol{\alpha} \equiv [\alpha_1, \ldots, \alpha_{366}]$. We chose the normalization $P_{GK}^1 \equiv 1$. The GK price and quantity levels have some good axiomatic properties including invariance to changes in the units of measurement.[22]

6.2 Linear Utility Function

Recall that we assumed utility maximizing behavior in Section 2 with the assumption of linear preferences. The first order conditions lead to the following *inverse demand estimating share equations*:[23]

$$s_{tn} = q_{tn} \alpha_n / (\boldsymbol{\alpha} \cdot \mathbf{q}^t); \quad t = 1, \ldots, 24; \; n = 1, \ldots, 366. \tag{6.3}$$

These share equations can be stacked up into a single estimating equation. The resulting nonlinear least squares minimization problem is:

$$\min_{\boldsymbol{\alpha}} \sum_{t=1}^{24} \sum_{n=1}^{366} \left\{ s_{tn} - \left[q_{tn} \alpha_n / (\boldsymbol{\alpha} \cdot \mathbf{q}^t) \right] \right\}^2. \tag{6.4}$$

If $\boldsymbol{\alpha}^*$ is a solution to (6.4), then so is $\lambda \boldsymbol{\alpha}^*$ for any positive scalar λ. Since product 162 is present in all 24 months, we impose the normalization $\alpha_{162} \equiv 1$.

[22] See Balk (1996) and Diewert (1999, 2023) on the test properties of various multilateral indexes.

[23] We work with inverse demand functions where period t prices are functions of period t expenditure e^t and the period t quantity vector $\mathbf{q}^t$. Thus we are estimating the direct utility function $f(\mathbf{q})$ instead of its dual unit cost function. When product n is not purchased in period t, the corresponding quantity q_{tn} is equal to 0. We have set $p_{tn} = 0$ by assumption for missing products. It is relatively easy to estimate inverse demand systems when there are missing products but more or less impossible to estimate traditional demand systems that estimate expenditure functions when there are missing products because the theoretically correct product prices are unobserved reservation prices.

The nonlinear regression option in Shazam was used to solve the minimization problem defined by (6.4). Denote the estimated α vector as $\boldsymbol{\alpha}^*$. Preliminary month t quantity and price levels were defined as $Q^{t*} \equiv \boldsymbol{\alpha}^* \cdot \mathbf{q}^t$ and $P^{t*} \equiv e^t/Q^{t*}$ for $t = 1, \ldots, 24$. These preliminary price levels were divided by P^{1*} in order to define the *Linear Utility Price indexes*, $P_{LU}^t \equiv P^{t*}/P^{1*}$ for $t = 1, \ldots, 24$. The R^2 for this regression was 0.9986.[24]

It is interesting to compare the P_{LU}^t price levels that are generated by this consumer demand model with the price levels P_{WTPD}^t that were generated by the Weighted Time Product Dummy Hedonic regression model estimated in Section 3.[25] Both models are based on the implicit or explicit assumption of linear purchaser preferences. For our data set, the two choices generated much the same indexes so the choice was not material.

6.3 CES Utility Function

A problem with the assumption of linear preferences is that it implies that the products are perfectly substitutable with each other; i.e., that the elasticity of substitution between products σ is equal to plus infinity. For our final index, we assume that purchasers have Constant Elasticity of Substitution preferences.[26]

The CES utility function, $f(\mathbf{q})$, is defined as follows:

$$f(\mathbf{q}) \equiv \left[\sum_{n=1}^{366} \alpha_n (q_n)^\kappa \right]^{1/\kappa} \tag{6.5}$$

where the α_n are positive parameters and the parameter κ satisfies the inequalities $0 < \kappa \leq 1$.[27] Note that if the parameter κ equals 1, then the CES utility function defined by (6.5) becomes the linear utility function.

We obtained estimates for the CES utility function by solving the following nonlinear

[24] We use the raw moment definition of R^2. Let $\mathbf{s}$ denote the vector of observed expenditure shares of dimension 8784, let $\mathbf{f}$ denote the vector of fitted shares and let $\mathbf{e}$ denote the error vector so that $\mathbf{s} = \mathbf{f} + \mathbf{e}$. Then $R^2 \equiv (\mathbf{s} \cdot \mathbf{f})^2/(\mathbf{s} \cdot \mathbf{s})(\mathbf{f} \cdot \mathbf{f})$.

[25] Note that both models generate normalized price levels that are invariant to changes in the units of measurement.

[26] This functional form was popularized by Arrow et al. (1961) in the context of production theory. For more on estimating CES utility functions, see Balk (1999) and de Haan and Krsinich (2024).

[27] We require that $\kappa \leq 1$ to ensure that the utility function is concave in the components of $\mathbf{q}$ and we require that $\kappa > 0$ in order to ensure that the utility function is well defined if any component of the $\mathbf{q}^t$ vector happens to be equal to 0.

least squares minimization problem:

$$\min_{\boldsymbol{\alpha},\kappa} \sum_{t=1}^{24} \sum_{n=1}^{366} \left\{ s_{tn} - \frac{\alpha_n (q_{tn})^\kappa}{\sum_{i=1}^{366} \alpha_i (q_{ti})^\kappa} \right\}^2 . \tag{6.6}$$

As was the case for the estimation of the linear utility function, we set $\alpha_{162} = 1$. The log likelihood for the linear utility function model was 58966.39. It took Shazam 317 iterations and 16.4 hours to converge. The final log likelihood for the new CES model was 58966.44. The estimate for κ was $\kappa^* \equiv 0.99962$, which is very close to 1, so the resulting CES estimated price and quantity indexes, P^t_{CES} and Q^t_{CES}, are very close to the corresponding indexes for the linear utility function model. The estimated elasticity of substitution was $\sigma^* \equiv 1/(1 - \kappa^*) = 2658.8$.

The five best indexes are compared in Table A-4 in the Appendix and plotted on Chart 4.

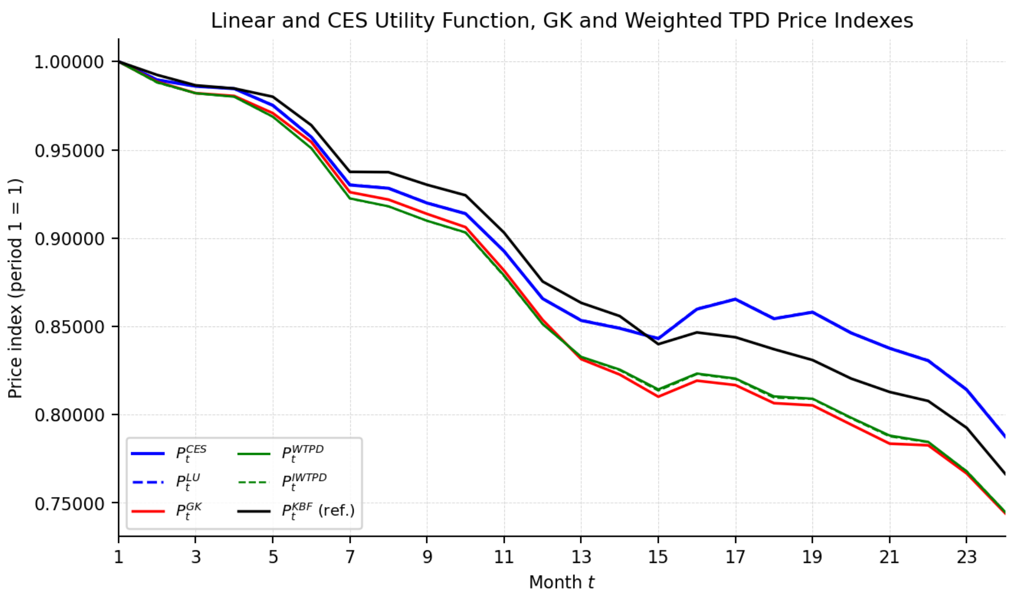

Figure 4: Linear and CES Utility Function, GK and Weighted Time Product Dummy Price Indexes

It can be seen that the CES and Linear Utility function price indexes, P^t_{CES} and P^t_{LU}, cannot be separately identified in the above Chart. The next 3 indexes, P^t_{GK}, P^t_{WTPD} and P^t_{IWTPD}, are very close to each other and lie below the first two indexes. The CES indexes are consistent with a more flexible functional form than the remaining 4 indexes, which are consistent with linear preferences.[28] But the estimated elasticity of substitution is so high that the CES index essentially collapses down to a linear utility function. Thus all 5 indexes should give much the same answer but it can be seen that the 5 indexes separate into two separate groups of indexes. For our particular data set, we cannot definitively single out any one of the above indexes as “best”. However, the fact that the CES model is consistent with both linear preferences and more general preferences (which do not assume that the products in scope are perfectly substitutable) cause us to lean towards this model.

7 Conclusion

We have calculated 17 separate indexes using our laptop data in the previous sections: P_A^t, P_{UV}^t, P_{TPD}^t, P_{WTPD}^t, P_{IWTPD}^t, P_{APTPD}^t, P_{WAPTPD}^t, P_S^t, P_{FCH}^t, P_{LCH}^t, P_{PCH}^t, P_{GEKS}^t, P_{EW}^t, P_{MEW}^t, P_{GK}^t, P_{LU}^t and P_{CES}^t.

The following tentative conclusions emerge from our study:

- When price and quantity data for the products in scope are available, it is best to use weighted hedonic regressions that take into account the economic importance of the products. We found substantial differences between our weighted and unweighted (or more accurately, equally weighted) hedonic regressions.

- There was a possible chain drift problem with all of our models that used chaining to link adjacent periods. The chain drift problem was not cured by the use of the multilateral predicted share method because most of the bilateral links chosen by the method were chain links.[29]

- Fixed Base indexes and GEKS and Rolling Window GEKS did not work well for our example due to the lack of matched products when the time period between

[28] The GK indexes are also consistent with Leontief (1947) preferences and the two Weighted Time Product Dummy indexes are also consistent with Cobb-Douglas (1928) preferences. But Leontief and Cobb-Douglas preferences are not plausible in our present context when there is tremendous product churn.

[29] The recent paper by Fox et al. (2025) found the same result for many product classes.

the base and current period grows.

- A possible solution to the chain drift problem for our example was provided by the use of the Expanding Window methodology explained in section 5. This method should work well for many product classes where substitution between the competing products is high and each product is available on the marketplace for a number of consecutive periods.

- Charts 2 and 4 show that different well established index number and consumer demand methods generate substantially different measures of price (and hence quantity) change.[30] Thus the method used to quality adjust products with a high degree of product churn matters. We could eliminate some methods as being unsatisfactory but we are not able to choose which method was "best" in a definitive manner between the 5 indexes that appeared on Chart 4.

Our research showed that many models may suffer from chain drift bias or from a lack of matching bias. There is a need to do additional research to see if the methods suggested in this paper to address these problems are robust.

[30] This "fact" helps to explain why inflation rates for high tech products differ so much across countries: a definitive method for quality adjustment has not yet emerged.

Appendix: Numerical Results

This appendix presents the complete monthly numerical results for all price indexes computed in sections 2–6, corresponding to the four charts in the main text, together with the matrix of Predicted Share dissimilarity measures discussed in section 4.

Table A-1: Time Product Dummy Price Indexes (Chart 1)

Table A-3 lists the seven indexes plotted on Chart 1: the simple average price index P_A^t, the unit value index P_{UV}^t, the unweighted and weighted sample-wide Time Product Dummy indexes P_{TPD}^t and P_{WTPD}^t, the Implicit Weighted TPD index P_{IWTPD}^t, and the unweighted and weighted Adjacent Period TPD indexes P_{APTPD}^t and P_{WAPTPD}^t. All indexes are normalised to unity at $t = 1$.

Table A-1: Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Price Indexes

Month	P_A^t	P_{UV}^t	P_{TPD}^t	P_{WTPD}^t	P_{IWTPD}^t	P_{APTPD}^t	P_{WAPTPD}^t
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.035	0.997	0.983	0.988	0.988	0.993	0.999
3	1.035	1.010	0.978	0.982	0.982	1.031	1.007
4	1.021	0.995	0.965	0.980	0.980	1.021	1.010
5	1.063	1.020	0.953	0.969	0.969	1.039	1.007
6	1.066	1.002	0.937	0.951	0.951	1.030	0.995
7	1.027	0.984	0.906	0.923	0.922	0.989	0.972
8	1.020	0.974	0.889	0.918	0.918	0.981	0.960
9	1.011	0.951	0.877	0.910	0.910	0.965	0.953
10	1.036	0.991	0.874	0.903	0.903	0.972	0.949
11	1.013	0.947	0.853	0.879	0.878	0.945	0.929
12	0.949	0.879	0.825	0.851	0.851	0.913	0.899
13	0.903	0.844	0.808	0.833	0.833	0.911	0.875
14	0.914	0.846	0.795	0.826	0.825	0.897	0.863
15	0.899	0.846	0.779	0.814	0.814	0.884	0.851
16	0.932	0.854	0.779	0.823	0.823	0.891	0.862
17	0.891	0.802	0.771	0.820	0.820	0.884	0.857
18	0.866	0.791	0.759	0.810	0.810	0.873	0.846
19	0.851	0.799	0.754	0.809	0.809	0.869	0.841
20	0.831	0.793	0.745	0.798	0.798	0.849	0.825
21	0.848	0.771	0.737	0.788	0.788	0.847	0.818
22	0.904	0.853	0.733	0.785	0.784	0.845	0.807
23	0.859	0.846	0.724	0.768	0.768	0.843	0.795
24	0.892	0.878	0.707	0.745	0.745	0.831	0.788
Mean	0.964	0.922	0.842	0.876	0.876	0.938	0.912

Table A-2: Similarity Linked and Traditional Price Indexes (Chart 2)

Table A-4 lists the eight indexes plotted on Chart 2: the Predicted Share similarity linked index P_S^t, the chained Fisher (P_{FCH}^t), Laspeyres (P_{LCH}^t) and Paasche (P_{PCH}^t) indexes, the sample-wide GEKS index P_{GEKS}^t, the Rolling Window GEKS with mean splice P_{RWGEKS}^t, and the Weighted Adjacent Period TPD index P_{WAPTPD}^t and Weighted TPD index P_{WTPD}^t (shown for reference).

Table A-2: Predicted Share Similarity Linked Index and Other Comparison Price Indexes

Month	P_S^t	P_{FCH}^t	P_{LCH}^t	P_{PCH}^t	P_{GEKS}^t	P_{RWGEKS}^t	P_{WAPTPD}^t	P_{WTPD}^t
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.993	0.993	0.995	0.991	0.990	0.988	0.999	0.988
3	0.985	0.985	0.985	0.984	1.005	0.983	1.007	0.982
4	0.983	0.983	0.983	0.983	1.016	0.982	1.010	0.980
5	0.979	0.972	0.970	0.975	0.994	0.978	1.007	0.969
6	0.968	0.962	0.959	0.965	0.984	0.966	0.995	0.951
7	0.948	0.941	0.939	0.944	0.954	0.938	0.972	0.923
8	0.935	0.927	0.924	0.930	0.946	0.928	0.960	0.918
9	0.925	0.918	0.912	0.923	0.933	0.917	0.953	0.910
10	0.926	0.918	0.905	0.932	0.930	0.919	0.949	0.903
11	0.894	0.889	0.872	0.907	0.906	0.888	0.929	0.879
12	0.862	0.857	0.841	0.873	0.880	0.856	0.899	0.851
13	0.828	0.824	0.811	0.836	0.859	0.822	0.875	0.833
14	0.817	0.813	0.803	0.823	0.853	0.816	0.863	0.826
15	0.798	0.794	0.784	0.805	0.836	0.797	0.851	0.814
16	0.797	0.792	0.781	0.804	0.841	0.800	0.862	0.823
17	0.789	0.785	0.773	0.796	0.846	0.798	0.857	0.820
18	0.780	0.776	0.765	0.786	0.824	0.779	0.846	0.810
19	0.768	0.764	0.755	0.773	0.841	0.773	0.841	0.809
20	0.753	0.749	0.740	0.757	0.848	0.759	0.825	0.798
21	0.743	0.739	0.733	0.746	0.836	0.753	0.818	0.788
22	0.734	0.730	0.724	0.736	0.835	0.746	0.807	0.785
23	0.715	0.711	0.707	0.716	0.794	0.728	0.795	0.768
24	0.693	0.690	0.689	0.690	0.760	0.710	0.788	0.745
Mean	0.860	0.855	0.847	0.862	0.898	0.862	0.912	0.876

Table A-3: Expanding Window Price Indexes (Chart 3)

Table A-3 reports the four indexes discussed in section 5: the Expanding Window index P_{EW}^t, the Modified Expanding Window index P_{MEW}^t, the Implicit Weighted Time Product Dummy index P_{IWTPD}^t, and the directly estimated Weighted Time Product Dummy index

P^t_{WTPD}. All four indexes are normalised to unity at $t = 1$.

The table shows that P^t_{EW} and P^t_{MEW} converge to the transitive target P^t_{IWTPD} by month 24 (all three equal 0.745), while P^t_{MEW} tracks P^t_{IWTPD} more closely during the first 12 months because it is initialised on a full retrospective window.

Table A-3: Expanding Window, Modified Expanding Window and Weighted Time Product Dummy Price Indexes

Month	$\boldsymbol{P^t_{EW}}$	$\boldsymbol{P^t_{MEW}}$	$\boldsymbol{P^t_{IWTPD}}$	$\boldsymbol{P^t_{WTPD}}$
1	1.000	1.000	1.000	1.000
2	0.994	0.988	0.988	0.988
3	0.985	0.979	0.982	0.982
4	0.984	0.976	0.980	0.980
5	0.985	0.966	0.969	0.969
6	0.973	0.950	0.951	0.951
7	0.943	0.928	0.922	0.923
8	0.934	0.920	0.918	0.918
9	0.921	0.913	0.910	0.910
10	0.908	0.906	0.903	0.903
11	0.884	0.884	0.878	0.879
12	0.855	0.855	0.851	0.851
13	0.822	0.822	0.833	0.833
14	0.811	0.811	0.825	0.826
15	0.797	0.797	0.814	0.814
16	0.806	0.806	0.823	0.823
17	0.811	0.811	0.820	0.820
18	0.797	0.797	0.810	0.810
19	0.799	0.799	0.809	0.809
20	0.792	0.792	0.798	0.798
21	0.784	0.784	0.788	0.788
22	0.786	0.786	0.784	0.785
23	0.766	0.766	0.768	0.768
24	0.745	0.745	0.745	0.745
Mean	0.870	0.866	0.870	0.870

Table A-4: Best Five Price Indexes from Preference-Based Models (Chart 4)

Table A-4 reports the five indexes discussed in section 6: the CES utility function index P^t_{CES}, the Linear Utility index P^t_{LU}, the Geary-Khamis index P^t_{GK}, the Weighted Time Product Dummy index P^t_{WTPD}, and the Implicit Weighted Time Product Dummy index P^t_{IWTPD}. All indexes are normalised to unity at $t = 1$ and are reported to five decimal places to reveal the small but systematic differences between them.

The table highlights two distinct clusters: the first (P^t_{CES} and P^t_{LU}) produces higher price levels throughout the sample, ending near 0.787–0.788 at month 24, while the second (P^t_{GK}, P^t_{WTPD}, P^t_{IWTPD}) ends near 0.744–0.745.

Table A-4: Linear and CES Utility Function Price Indexes, Geary Khamis, Weighted TPD and Implicit Weighted Time Product Dummy Price Indexes

Month t	$\boldsymbol{P^t_{CES}}$	$\boldsymbol{P^t_{LU}}$	$\boldsymbol{P^t_{GK}}$	$\boldsymbol{P^t_{WTPD}}$	$\boldsymbol{P^t_{IWTPD}}$
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.98960	0.98966	0.98855	0.98828	0.98800
3	0.98599	0.98608	0.98209	0.98205	0.98203
4	0.98464	0.98465	0.98050	0.98006	0.98006
5	0.97518	0.97525	0.97078	0.96878	0.96867
6	0.95721	0.95726	0.95443	0.95087	0.95101
7	0.93006	0.93020	0.92600	0.92250	0.92232
8	0.92822	0.92829	0.92181	0.91801	0.91792
9	0.91986	0.91992	0.91371	0.90983	0.90975
10	0.91384	0.91388	0.90620	0.90323	0.90303
11	0.89257	0.89258	0.88160	0.87881	0.87834
12	0.86571	0.86567	0.85367	0.85129	0.85132
13	0.85332	0.85345	0.83144	0.83276	0.83257
14	0.84888	0.84905	0.82270	0.82554	0.82512
15	0.84305	0.84323	0.81018	0.81431	0.81357
16	0.85972	0.85988	0.81924	0.82328	0.82289
17	0.86538	0.86548	0.81675	0.82048	0.82014
18	0.85432	0.85444	0.80651	0.81037	0.80959
19	0.85803	0.85810	0.80527	0.80906	0.80882
20	0.84628	0.84634	0.79437	0.79830	0.79788
21	0.83758	0.83764	0.78358	0.78818	0.78764
22	0.83056	0.83070	0.78265	0.78460	0.78436
23	0.81405	0.81421	0.76665	0.76781	0.76800
24	0.78754	0.78771	0.74412	0.74478	0.74489
Mean	0.89340	0.89349	0.86928	0.86972	0.86950

Table A-5: Predicted Share Dissimilarity Measures

Table A-5 lists the 24 × 24 matrix of Predicted Share measures of relative price dissimilarity defined by equation (4.5). Entry (r, t) measures the dissimilarity between months r and t. The matrix is symmetric and has zeros on the diagonal. Small values indicate similar relative price structures; large values indicate substantial dissimilarity. The similarity-linked Predicted Share index P^t_S reported in Table A-2 is constructed by linking

Fisher bilateral price indexes between the pair of periods with the smallest dissimilarity measure.

Table A-5: Predicted Share Dissimilarity Measures $\Delta(\mathbf{p}^r, \mathbf{p}^t, \mathbf{q}^r, \mathbf{q}^t) \times 10^2$ (row $= r$, column $= t$)

	Month t																							
r	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1	0.00	1.03	0.88	1.70	3.12	4.92	5.14	5.06	7.19	6.43	8.76	10.09	13.96	14.12	14.87	17.84	29.95	37.98	39.37	60.77	58.92	84.98	86.46	101.32
2	1.03	0.00	0.07	0.92	1.46	2.57	2.68	3.25	4.10	4.48	5.46	5.54	8.32	9.35	10.13	11.58	23.56	29.93	34.28	50.73	50.08	67.05	65.71	85.55
3	0.88	0.07	0.00	0.46	0.57	1.19	1.63	1.68	2.29	2.36	3.19	3.40	4.97	5.68	6.20	7.99	14.80	17.19	28.43	32.55	28.37	44.50	49.14	61.26
4	1.70	0.92	0.46	0.00	1.16	1.49	2.10	1.96	2.67	2.68	4.14	4.59	5.00	5.45	5.66	7.67	12.92	14.42	25.45	25.34	22.33	37.99	45.68	45.93
5	3.12	1.46	0.57	1.16	0.00	0.05	0.79	0.30	0.74	0.71	1.73	2.15	2.85	3.47	4.05	5.11	9.29	8.52	15.47	17.32	15.53	23.17	36.29	31.82
6	4.92	2.57	1.19	1.49	0.05	0.00	0.75	0.27	0.66	0.59	1.64	2.07	2.76	3.35	3.97	4.83	8.65	7.68	15.49	16.64	14.73	22.16	37.30	30.71
7	5.14	2.68	1.63	2.10	0.79	0.75	0.00	0.45	0.44	0.57	0.67	0.75	2.40	3.20	3.68	4.57	9.26	8.29	15.83	17.24	18.49	24.61	42.68	35.39
8	5.06	3.25	1.68	1.96	0.30	0.27	0.45	0.00	0.02	0.13	0.07	0.12	1.60	2.20	2.66	3.45	7.58	6.67	13.81	15.25	16.59	24.65	40.61	26.08
9	7.19	4.10	2.29	2.67	0.74	0.66	0.44	0.02	0.00	0.09	0.02	0.05	1.44	2.40	2.95	3.74	7.63	6.87	13.92	15.32	16.57	24.42	40.61	26.26
10	6.43	4.48	2.36	2.68	0.71	0.59	0.57	0.13	0.09	0.00	0.07	0.39	1.74	2.30	2.89	3.67	7.75	6.65	14.09	15.43	16.77	24.63	41.02	26.12
11	8.76	5.46	3.19	4.14	1.73	1.64	0.67	0.07	0.02	0.07	0.00	0.02	1.33	1.85	2.39	3.20	7.44	6.82	13.44	15.71	17.11	24.57	41.65	28.16
12	10.09	5.54	3.40	4.59	2.15	2.07	0.75	0.12	0.05	0.39	0.02	0.00	1.32	1.81	2.37	3.42	8.60	8.21	14.29	18.50	19.64	28.96	46.28	32.49
13	13.96	8.32	4.97	5.00	2.85	2.76	2.40	1.60	1.44	1.74	1.33	1.32	0.00	0.35	0.32	0.57	1.84	2.30	3.55	3.80	4.43	8.42	10.22	9.37
14	14.12	9.35	5.68	5.45	3.47	3.35	3.20	2.20	2.40	2.30	1.85	1.81	0.35	0.00	0.06	0.31	1.11	1.70	2.48	2.54	2.99	6.56	7.67	7.62
15	14.87	10.13	6.20	5.66	4.05	3.97	3.68	2.66	2.95	2.89	2.39	2.37	0.32	0.06	0.00	0.03	0.39	0.72	1.12	1.01	1.48	4.86	5.50	5.67
16	17.84	11.58	7.99	7.67	5.11	4.83	4.57	3.45	3.74	3.67	3.20	3.42	0.57	0.31	0.03	0.00	0.14	0.35	0.44	0.45	0.64	4.07	4.34	4.58
17	29.95	23.56	14.80	12.92	9.29	8.65	9.26	7.58	7.63	7.75	7.44	8.60	1.84	1.11	0.39	0.14	0.00	0.20	0.25	0.25	0.36	3.91	4.12	4.38
18	37.98	29.93	17.19	14.42	8.52	7.68	8.29	6.67	6.87	6.65	6.82	8.21	2.30	1.70	0.72	0.35	0.20	0.00	0.12	0.31	0.19	3.59	3.58	3.96
19	39.37	34.28	28.43	25.45	15.47	15.49	15.83	13.81	13.92	14.09	13.44	14.29	3.55	2.48	1.12	0.44	0.25	0.12	0.00	0.06	0.10	3.49	3.32	3.67
20	60.77	50.73	32.55	25.34	17.32	16.64	17.24	15.25	15.32	15.43	15.71	18.50	3.80	2.54	1.01	0.45	0.25	0.31	0.06	0.00	0.06	3.41	3.36	3.70
21	58.92	50.08	28.37	22.33	15.53	14.73	18.49	16.59	16.57	16.77	17.11	19.64	4.43	2.99	1.48	0.64	0.36	0.19	0.10	0.06	0.00	3.30	3.13	3.56
22	84.98	67.05	44.50	37.99	23.17	22.16	24.61	24.65	24.42	24.63	24.57	28.96	8.42	6.56	4.86	4.07	3.91	3.59	3.49	3.41	3.30	0.00	0.09	0.43
23	86.46	65.71	49.14	45.68	36.29	37.30	42.68	40.61	40.61	41.02	41.65	46.28	10.22	7.67	5.50	4.34	4.12	3.58	3.32	3.36	3.13	0.09	0.00	0.13
24	101.32	85.55	61.26	45.93	31.82	30.71	35.39	26.08	26.26	26.12	28.16	32.49	9.37	7.62	5.67	4.58	4.38	3.96	3.67	3.70	3.56	0.43	0.13	0.00

1. Conflict of Interest

The authors declare that they have no conflicts of interest. Neither author has any financial or personal relationships with other people or organisations that could inappropriately influence (bias) their work.

2. Funding

This research received financial support from the Japan Society for the Promotion of Science (JSPS KAKENHI Grant 24H00012).

3. Data and Code Availability

The data and programme code used in this study will be made publicly available upon acceptance of the paper. Each dataset and the accompanying replication code will be assigned a Digital Object Identifier (DOI) to ensure permanent, citable access for readers wishing to verify or extend the empirical results.

4. Authors' Contributions

W. Erwin Diewert supervised the entire research project, led the development of the theoretical framework, and oversaw the empirical analysis throughout. He also had overall responsibility for drafting and revising the manuscript. Chihiro Shimizu was responsible for data collection and preparation, and undertook the full empirical analysis, including the construction and computation of all price indexes.

5. Ethics Approval

This study uses only publicly available, aggregated sales data on laptop computers (prices and quantities) provided by a private data vendor. No human participants, personal data, or biological material were involved in the research. Accordingly, approval by an ethics committee or institutional review board was not required.

6. Use of Artificial Intelligence Tools

No artificial intelligence tools, including large language models, were used in the preparation or writing of this manuscript.

References

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.

Balk, B.M. (1996), "A Comparison of Ten Methods for Multilateral International Price and Volume Comparisons", *Journal of Official Statistics* 12, 199–222.

Balk, B.M. (1999), "On Curing the CPI's Substitution and New Goods Bias", Paper presented at the Fifth Meeting of the Ottawa Group, Reykjavik, Iceland.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), "Multilateral Comparisons of Output, Input, and Productivity using Superlative Index Numbers", *Economic Journal* 92, 73–86.

Chessa, A.G. (2016), "A New Methodology for Processing Scanner Data in the Dutch CPI", *EURONA* 1, 49–69.

Chessa, A.G. (2021), "Extension of Multilateral Index Series Over Time: Analysis and Comparison of Methods", paper written for the 2021 Meeting of the Group of Experts on Consumer Price Indices. Geneva, 7 May 2021. https://unece.org/sites/default/files/2021-05/Session_1_Netherlands_Paper.pdf

Cobb, C. and P.H. Douglas (1928), "A Theory of Production", *American Economic Review* 18, 139–165.

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. (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. (2010), "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Repricing Methods", *Jahrbücher für Nationökonomie und Statistik* 230, 772–791.

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, 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.

de Haan, J. and F. Krsinich (2024), “Product Churn and the GEKS-Törnqvist Price Index: The Feenstra Adjustment”, paper presented at the 18th Meeting of the Ottawa Group, Ottawa, Canada.

de Haan, J., R. Hendriks and M. Scholz (2021), “Price Measurement using Scanner Data: Time Product Dummy versus Time Dummy Hedonic Indexes”, *Review of Income and Wealth* 67:2, 394–417.

Diewert, W.E. (1999), “Axiomatic and Economic Approaches to International Comparisons”, in *International and Interarea Comparisons of Income, Output and Prices*, W.E. Diewert (ed.), 13–87. Amsterdam: Elsevier.

Diewert, W.E. (2003a), “Hedonic Regressions: A Consumer Theory Approach”, in *Scanner Data and Price Indexes*, R.C. Feenstra and M.D. Shapiro (eds.), 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. (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. (2023), “Scanner Data, Elementary Price Indexes and the Chain Drift

Problem", pp. 445–606 in *Advances in Economic Measurement*, D. Chotikapanich, A.N. Rambaldi and N. Rhode (eds.), Singapore: Palgrave Macmillan.

Diewert, W.E. (2025), "Quality Adjustment", Chapter 8 in *Consumer Price Index Manual 2025*, Washington D.C.: International Monetary Fund. https://www.imf.org/en/Data/Statistics/cpi-manual

Diewert, W.E. and K.J. Fox (2022), "Substitution Bias in Multilateral Methods for CPI Construction", *Journal of Business and Economic Statistics* 40:1, 355–369.

Diewert, W.E. and C. Shimizu (2026), "Scanner Data, Product Churn and Quality Adjustment", NBER Working Paper 34897, Cambridge MA: National Bureau of Economic Research.

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

Fox, K.J., P. Levell and M. O'Connell (2025), "Inflation Measurement with High Frequency Data", Working Paper 23/29, Institute for Fiscal Studies, London, U.K., forthcoming in the *Journal of Business and Economic Statistics*.

Geary, R.G. (1958), "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society Series A* 121, 97–99.

Gini, C. (1931), "On the Circular Test of Index Numbers", *Metron* 9:9, 3–24.

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. Pólya (1934), *Inequalities*, Cambridge: Cambridge University Press.

Hill, R.J. (1997), "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities", *Review of Income and Wealth* 43:1, 49–69.

Hill, R.J. (2001), "Measuring Inflation and Growth Using Spanning Trees", *International Economic Review* 42, 167–185.

Hill, R.J. (2004), "Constructing Price Indexes Across Space and Time: The Case of the European Union", *American Economic Review* 94, 1379–1410.

Inklaar, R. and W.E. Diewert (2016), "Measuring Industry Productivity and Cross-Country Convergence", *Journal of Econometrics* 191, 426–433.

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.

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.

Khamis, S.H. (1970), "Properties and Conditions for the Existence of a New Type of Index Number", *Sankhya B* 32, 81–98.

Krsinich, F. (2016), "The FEWS Index: Fixed Effects with a Window Splice", *Journal of Official Statistics* 32:2, 375–404.

Leontief, W. (1947), "Introduction to a Theory of the Internal Structure of Functional Relationships", *Econometrica* 15:4, 361–373.

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

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

Shankar, S. and G. Hajargasht (2017), "A Stochastic Approach to Index Numbers for Regional Price Comparisons", *Review of Income and Wealth* 63, S153–S175.

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

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

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

Walsh, C.M. (1921), “Discussion”, *Journal of the American Statistical Association* 17, 537–544.