Scanner Data, Elementary Price Indexes and the Chain Drift Problem

Erwin Diewert,[1]
Discussion Paper 18-06,
Vancouver School of Economics,
University of British Columbia,
Vancouver, B.C., Canada, V6T 1L4.

Revised October 7, 2018.

Abstract

Statistical agencies increasingly are able to collect detailed price and quantity information from retailers on sales of consumer products. Thus elementary price indexes (which are indexes constructed at the first stage of aggregation for closely related products) can now be constructed using this price and quantity information, whereas previously, statistical agencies had to construct elementary indexes using just retail outlet collected information on prices alone. Thus superlative indexes can now be constructed at the elementary level, which in theory, should lead to more accurate Consumer Price Indexes. However, retailers frequently sell products at heavily discounted prices, which lead to large increases in purchases of these products. This volatility in prices and quantities will generally lead to a chain drift problem; i.e., when prices return to their "normal" levels, quantities purchased are frequently below their "normal" levels and this leads to a downward drift in a superlative price index. The paper addresses this problem and looks at the likely bias in various index number formulae that are commonly used. The bias estimates are illustrated using some scanner data on the sales of frozen juice products that are available online.

Keywords

Jevons, Dutot, Carli, Unit Value, Laspeyres, Paasche, Constant Elasticity of Substitution (CES), Fisher and Törnqvist price indexes, superlative indexes, multilateral indexes, GEKS, CCDI, Geary-Khamis, Similarity Linked and Time Product Dummy indexes, elementary indexes, scanner data, bias estimates.

JEL Classification Numbers

C43, C81, E31.

---

[1] W. Erwin Diewert: Vancouver School of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and the School of Economics, UNSW Sydney, NSW 2052, Australia (erwin.diewert@ubc.ca). The author thanks Corinne Becker-Vermeulen, Jan de Haan, Claude Lamboray and Mick Silver for helpful comments and gratefully acknowledges the financial support of a Trans Atlantic Platform Digging into Data grant. The project title is: Online Prices for Computing Standards of Living Across Countries (OPSLAC).

Table of Contents

1. Introduction

Statistical agencies increasingly are approaching retail chains and asking them to provide information on weekly values and quantities sold of scanner coded products and in many countries, this information is being provided to the relevant agency. This detailed weekly information on the value and quantity of sales by product means that it is possible to construct index numbers in real time that require product information on prices and quantities, such as the Fisher (1922) ideal index. The *Consumer Price Index Manual*[2] that was published in 2004 recommended that chained superlative index numbers,[3] such as the Fisher index, be used in this situation when detailed price and quantity information is available in real time. However, experience with scanner data has shown that the use of chained superlative indexes leads to *chain drift*; i.e., a growing divergence between the chained index and its counterpart fixed base index.[4]

There are at least three possible real time solutions to the chain drift problem:

- Use a fixed base index;
- Use a multilateral index;[5]
- Use annual weights for a past year.

There are two problems with the first solution: (i) the results depend asymmetrically on the choice of the base period and (ii) with new and disappearing products, the base period prices and quantities may lose their representativeness; i.e., over long periods of time, matching products becomes very difficult.[6] A problem with the second solution is that as an extra period of data becomes available, the indexes have to be recomputed.[7] The

[2] See ILO, IMF, OECD, Eurostat, UN and the World Bank (2004).

[3] See Diewert (1976) on the concept of a superlative index. Basically, a superlative price index allows for an arbitrary pattern of substitution effects between products.

[4] Fisher (1922; 293) realized that the chained Carli, Laspeyres and Young indexes were subject to upward chain drift but for his empirical example, there was no evidence of chain drift for the Fisher formula. However, Persons (1921) came up with an actual empirical example where the Fisher index exhibited substantial downward chain drift. Frisch (1936; 9) seems to have been the first to use the term “chain drift”. Both Frisch (1936; 8-9) and Persons (1928; 100-105) discussed and analyzed the chain drift problem.

[5] The use of multilateral indexes in the time series context dates back to Persons (1921) and Fisher (1922; 297-308), Gini (1931) and Balk (1980) (1981). Fisher (1922; 305) suggested taking the arithmetic average of the Fisher “star” indexes whereas Gini suggested taking the geometric mean of the star indexes.

[6] Persons (1928; 99-100) has an excellent discussion on the difficulties of matching products over time.

[7] This is not a major problem. A solution to this problem is to use a rolling window of observations and use the results of the current window to update the index to the current period. This methodology was suggested by Ivancic, Diewert and Fox (2009) (2011) and is being used by the Australian Bureau of Statistics (2016). Ivancic, Diewert and Fox (2011) suggested that the movement of the indexes for the last two periods in the new window be linked to the last index value generated by the previous window. However Krsinich (2016) in a slightly different context suggested that the movement of the indexes generated by the new window over the entire new window period be linked to the previous window index value for the second period in the previous window. Krsinich called this a *window splice* as opposed to the IDF *movement splice*. De Haan (2015; 27) suggested that perhaps the linking period should be in the middle of the old window which the Australian Bureau of Statistics (2016; 12) terms a *half splice*. Ivancic, Diewert and Fox (2010) also suggested that the *average* of all links for the last period in the new window to

problem with the third possible solution is that the use of annual weights will inevitably result in some substitution bias. In any case, in this study, we will look at many of the commonly used fixed base indexes as well as six multilateral indexes and two indexes that use annual weights. We will develop approximate and exact relationships between these indexes and indicate likely differences (or "biases") between the indexes.

Statistical agencies are also using web-scraping to collect large number of prices as a substitute for selective sampling of prices at the first stage of aggregation. Thus it is of interest to look at elementary indexes that depend only on prices, such as the Carli (1804), Dutot (1838) and Jevons (1865) indexes, and compare these indexes to superlative indexes; i.e., under what conditions will these indexes adequately approximate a superlative index.[8]

The two superlative indexes that we will consider in this study are the Fisher (1922) and the Törnqvist[9] indexes. The reasons for singling out these two indexes as preferred bilateral index number formulae are as follows: (i) both indexes can be given a strong justification from the viewpoint of the economic approach to index number theory;[10] (ii) the Fisher index emerges as probably being the "best" index from the viewpoint of the axiomatic or test approach to index number theory;[11] (iii) the Törnqvist index has a strong justification from the viewpoint of the stochastic approach to index number theory.[12] Thus there are strong cases for the use of these two indexes when making comparisons of prices between two periods when detailed price and quantity data are available.

When comparing two indexes, two methods for making the comparisons will be used: (i) use second order Taylor series approximations to the index differences; (ii) the difference between two indexes can frequently be written as a covariance and it is possible in many cases to determine the likely sign of the covariance.[13]

When looking a scanner data from a retail outlet (or price and quantity data from a firm that uses dynamic pricing to price its products or services[14]), a fact emerges: if a product or a service is offered at a highly discounted price (i.e., it goes on sale), then the quantity sold of the product can increase by a very large amount. This empirical observation will allow us to make reasonable guesses about the signs of various covariances that express the difference between two indexes. If we are aggregating products that are close

---

the observations in the old window could be used as the linking factor. Diewert and Fox (2017) look at the alternative methods for linking.

[8] We will also look at the properties of the CES price index with equal weights.

[9] The usual reference is Törnqvist (1936) but the index formula did not actually appear in this paper. It did appear explicitly in Törnqvist and Törnqvist (1937). It was listed as one of Fisher's (1922) many indexes: namely number 123. It was explicitly recommended as one of his top five ideal indexes by Warren Persons (1928; 86) so it probably should be called the Persons index.

[10] The economic approach to index number theory is due to Konüs (1924). See Diewert (1976) for justifications for the use of these two indexes from the viewpoint of the economic approach to index number theory.

[11] See Diewert (1992) or Chapter 16 of the *Consumer Price Index Manual*.

[12] See Theil (1967; 136-137) or Chapter 16 of the *Consumer Price Index Manual*.

[13] This second method for making comparisons can be traced back to Bortkiewicz (1923).

[14] Airlines and hotels are increasingly using dynamic pricing; i.e., they change prices frequently.

substitutes for each other, then a heavily discounted price may not only increase the sales of the product but it may also increase the *share* of the sales in the list of products or services that are in scope for the index.[15] It turns out that the behavior of shares in response to discounted prices does make a difference in analyzing the differences between various indexes: in the context of highly substitutable products, a heavily discounted price will probably increase the market share of the product but if the products are weak substitutes (which is typically the case at higher levels of aggregation), then a discounted price will typically increase sales of the product but not increase its market share. These two cases (strong or weak substitutes) will play an important role in our analysis.

Sections 2 and 3 look at relationships between the fixed base and chained Carli, Dutot, Jevons and CES elementary indexes that do not use share or quantity information.

Section 4 looks at the relationships between the Laspeyres, Paasche, Geometric Laspeyres, Geometric Paasche, Fisher and Törnqvist price indexes. Section 5 investigates how close the unweighted Jevons index is to the Geometric Laspeyres, Geometric Paasche $P_{GP}{}^t$ and Törnqvist $P_T{}^t$ price indexes.

Section 6 develops some relationships between the Törnqvist index and geometric indexes that use average annual shares as weights.

Section 7 looks at the differences between fixed base and chained Törnqvist indexes.

Multilateral indexes make an appearance in section 8: the fixed base Törnqvist index is compared to the GEKS and GEKS-Törnqvist (or CCDI) multilateral indexes.

Sections 9 and 10 compare Unit Value and Quality Adjusted Unit Value indexes to the Fisher index while section 11 compares the Lowe index to the Fisher index.

Sections 12-14 look at various additional multilateral indexes: the Geary Khamis in section 12, the Weighted Time Product Dummy index in section 13 and Similarity Linked price indexes in section 14.

The Appendix evaluates all of the above indexes for a grocery store scanner data set that is publically available. However, the data set had a number of missing prices and quantities. Some of these missing prices may be due to lack of sales or shortages of inventory. In addition, how should the introduction of new products and the disappearance of (possibly) obsolete products be treated in the context of forming a consumer price index? Hicks (1940; 140) suggested a general approach to this measurement problem in the context of the economic approach to index number theory. His approach was to apply normal index number theory but estimate (or guess at) hypothetical prices that would induce utility maximizing purchasers of a related group of products to demand 0 units of unavailable products. With these virtual (or reservation or imputed) prices in hand, one can just apply normal index number theory using the

[15] In the remainder of this study, we will speak of products but the same analysis applies to services.

augmented price data and the observed quantity data. In our empirical example in the Appendix, we will use the scanner data that was used in Diewert and Feenstra (2017) for frozen juice products for a Dominick's store in Chicago for 3 years. This data set had 20 observations where $q_{tn} = 0$. For these 0 quantity observations, Diewert and Feenstra estimated positive Hicksian reservation prices for these missing price observations and these imputed prices are used in our empirical example in the Appendix. However, it is possible to use alternative positive reservation prices that do not rely on econometrics. In section 15 below, we will discuss one of these alternative methods for constructing reservation prices for observations when sales of a product are zero. We used an inflation adjusted carry forward and backward methodology to construct alternative reservation prices for the missing prices and compared selected indexes using the new reservation prices to the corresponding indexes using the econometrically estimated reservation prices. The resulting differences in our "best" index numbers for our example data set are listed in section A.6 of the Appendix.

The Appendix lists the Dominick's data along with the estimated reservation prices. The Appendix also has tables and charts of the various index number formulae that are discussed in the main text of the study.

Section 16 concludes.

2. Comparing CES Price Levels and Indexes

In this section, we will begin our analysis by considering alternative methods by which the prices for N related products could be aggregated into a "representative" or aggregate price for the products for a given period.

We introduce some notation that will be used in the rest of the paper. We suppose that we have collected price and quantity data from a retail outlet on N closely related products for T time periods.[16] Typically, a time period is a month. Denote the price of product n in period t as p_{tn} and the corresponding quantity during period t as q_{tn} for $n = 1,...,N$ and $t = 1,...,T$. Usually, p_{tn} will be the period t *unit value price* for product n in period t; i.e., $p_{tn} \equiv v_{tn}/q_{tn}$ where v_{tn} is the total value of product n that is sold during period t and q_{tn} is the total quantity of product n that is sold during period t. We assume that $q_{tn} \geq 0$ and $p_{tn} > 0$ for all t and n.[17] The restriction that all products have positive prices associated with them is a necessary one for much of our analysis since many popular index numbers are constructed using logarithms of prices and the logarithm of a zero price is not well defined. However, our analysis does allow for possible 0 quantities and values sold during periods in the sample. Denote the period t strictly positive *price vectors* and nonnegative and non zero *quantity vectors* as $p^t \equiv [p_{t1},...,p_{tN}] >> 0_N$ and $q^t \equiv [q_{t1},...,q_{tN}] >$

[16] The T periods can be regarded as a window of observations, followed by another window of length T which has dropped the first period from the window and added the data of period T+1 to the window. The literature on how to link the results of one window to the next window is discussed in Diewert and Fox (2017). We will not discuss this linking problem in the present study.

[17] In the case where $q_{tn} = 0$, then $v_{tn} = 0$ as well and hence $p_{tn} \equiv v_{tn}/q_{tn}$ is not well defined in this case. In the case where $q_{tn} = 0$, we will assume that p_{tn} is a positive imputed price.

0_N respectively for t = 1,...,T where 0_N is an N dimensional vector of zeros. The inner product of the vectors p^t and q^t is denoted by $p^t \cdot q^t \equiv \Sigma_{n=1}^N p_{tn}q_{tn} > 0$. Define the period t sales (or expenditure) share for product n as $s_{tn} \equiv p_{tn}q_{tn}/p^t \cdot q^t$ for n = 1,...,N and t = 1,...,T. The *period t sales share vector* is defined as $s^t \equiv [s_{t1},...,s_{tN}] > 0_N$ for t = 1,...,T.

In most applications, the N products are closely related and they have common units of measurement (by weight, or by volume or by "standard" package size). In this context, it is useful to define the period t "real" share for product n of total product sales, $S_{tn} \equiv q_{tn}/1_N \cdot q^t$ for n = 1,...,N and t = 1,...,T where 1_N is an N dimensional vector of ones. Denote the *period t real share vector* as $S^t \equiv [S_{t1},...,S_{tN}]$ for t = 1,...,T.

Define the generic *product weighting vector* as $\alpha \equiv [\alpha_1,...,\alpha_N]$. We assume that α has strictly positive components which sum to one; i.e., we assume that α satisfies:

(1) $\alpha \cdot 1_N = 1$; $\alpha >> 0_N$.

Let $p \equiv [p_1,...,p_N] >> 0_N$ be a positive price vector. The corresponding *mean of order r of the prices p (with weights α)* or *CES price level*, $m_{r,\alpha}(p)$ is defined as follows:[18]

(2) $m_{r,\alpha}(p) \equiv [\Sigma_{n=1}^N \alpha_n p_n^{\,r}]^{1/r}$; $r \neq 0$;
$\equiv \prod_{n=1}^N (p_n)^{\alpha_n}$; $r = 0$.

It is useful to have a special notation for $m_{r,\alpha}(p)$ when r = 1:

(3) $p_\alpha \equiv \Sigma_{n=1}^N \alpha_n p_n = \alpha \cdot p$.

Thus p_α is an α weighted arithmetic mean of the prices $p_1,p_2,...,p_N$ and it can be interpreted as a *weighted Dutot price level*.[19]

From Schlömilch's (1858) Inequality,[20] we know that $m_{r,\alpha}(p) \geq m_{s,\alpha}(p)$ if $r \geq s$ and $m_{r,\alpha}(p) \leq m_{s,\alpha}(p)$ if $r \leq s$. However, we do not know how big the gaps are between these price levels for different r and s. When r = 0, $m_{0,\alpha}(p)$ becomes a weighted geometric mean or a *weighted Jevons* (1865) or *Cobb-Douglas price level* and it is of interest to know how much higher the weighted Dutot price level is than the corresponding weighted Jevons price level. Proposition 1 below provides an approximation to the gap between $m_{r,\alpha}(p)$ and $m_{1,\alpha}(p)$ for any r, including r = 0.

[18] Hardy, Littlewood and Polya (1934; 12-13) refer to this family of means or averages as *elementary weighted mean values* and study their properties in great detail. The function $m_{r,\alpha}(p)$ can also be interpreted as a *Constant Elasticity of Substitution (CES) unit cost function* if $r \leq 1$. The corresponding utility or production function was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961). For additional material on CES functions, see Feenstra (1994) and Diewert and Feenstra (2017).
[19] The ordinary Dutot (1738) price level for the period t prices p^t is defined as $p_D^{\,t} \equiv (1/N)\Sigma_{n=1}^N p_{tn}$. Thus it is equal to $m_{1,\alpha}(p^t)$ where $\alpha = (1/N)1_N$.
[20] See Hardy, Littlewood and Polya (1934; 26) for a proof of this result.

Define the α *weighted variance* of $p/p_\alpha \equiv [p_1/p_\alpha,...,p_N/p_\alpha]$ where p_α is defined by (3) as follows:[21]

$$(4)\ \mathrm{Var}_\alpha(p/p_\alpha) \equiv \Sigma_{n=1}^N \alpha_n[(p_n/p_\alpha) - 1]^2 .$$

Proposition 1: Let $p >> 0_N$, $\alpha >> 0_N$ and $\alpha\cdot 1_N = 1$. Then $m_{r,\alpha}(p)/m_{1,\alpha}(p)$ is approximately equal to the following expression for any r:

$$(5)\ m_{r,\alpha}(p)/m_{1,\alpha}(p) \approx 1 + (½)(r-1)\mathrm{Var}_\alpha(p/p_\alpha)$$

where $\mathrm{Var}_\alpha(p/p_\alpha)$ is defined by (4). The expression on the right hand side of (5) uses a second order Taylor series approximation to $m_{r,\alpha}(p)$ around the equal price point $p = p_\alpha 1_N$ where p_α is defined by (3).[22]

Proof: Straightforward calculations show that the level, vector of first order partial derivatives and matrix of second order partial derivatives of $m_{r,\alpha}(p)$ evaluated at the equal price point $p = p_\alpha 1_N$ are equal to the following expressions: $m_{r,\alpha}(p_\alpha 1_N) = p_\alpha \equiv \alpha\cdot p$; $\nabla_p m_{r,\alpha}(p_\alpha 1_N) = \alpha$; $\nabla^2_{pp} m_{r,\alpha}(p_\alpha 1_N) = (p_\alpha)^{-1}(r-1)(\hat{\alpha} - \alpha\alpha^T)$ where $\hat{\alpha}$ is a diagonal N by N matrix with the elements of the column vector α running down the main diagonal and α^T is the transpose of the column vector α. Thus $\alpha\alpha^T$ is a rank one N by N matrix.

Thus the second order Taylor series approximation to $m_{r,\alpha}(p)$ around the point $p = p_\alpha 1_N$ is given by the following expression:

$$\begin{aligned}(6)\ m_{r,\alpha}(p) &\approx p_\alpha + \alpha\cdot(p - p_\alpha 1_N) + (½)(p - p_\alpha 1_N)^T(p_\alpha)^{-1}(r-1)(\hat{\alpha} - \alpha\alpha^T)(p - p_\alpha 1_N) \\ &= p_\alpha + (½)(p_\alpha)^{-1}(r-1)(p - p_\alpha 1_N)^T(p_\alpha)^{-1}(\hat{\alpha} - \alpha\alpha^T)(p - p_\alpha 1_N) \quad \text{using (1) and (3)} \\ &= p_\alpha[1 + (½)(r-1)(p_\alpha)^{-2}(p - p_\alpha 1_N)^T(\hat{\alpha} - \alpha\alpha^T)(p - p_\alpha 1_N)] \\ &= m_{1,\alpha}(p)[1 + (½)(r-1)\mathrm{Var}_\alpha(p/p_\alpha)] \quad \text{using (2), (3) and (4).}\end{aligned}$$

Q.E.D.

The approximation (6) also holds if r = 0. In this case, (6) becomes the following approximation:[23]

$$\begin{aligned}(7)\ m_{0,\alpha}(p) &\equiv \prod_{n=1}^N (p_n)^{\alpha_n} \\ &\approx m_{1,\alpha}(p)[1 - (½)\mathrm{Var}_\alpha(p/p_\alpha)] \\ &= m_{1,\alpha}(p)\{1 - (½)\Sigma_{n=1}^N \alpha_n[(p_n/p_\alpha) - 1]^2\} \\ &= [\Sigma_{n=1}^N \alpha_n p_n]\{1 - (½)\Sigma_{n=1}^N \alpha_n[(p_n/p_\alpha) - 1]^2\}\end{aligned}$$

---

[21] Note that the α weighted mean of p/p_α is equal to $\Sigma_{n=1}^N \alpha_n p_n/p_\alpha = 1$. Thus (4) defines the corresponding weighted variance.

[22] For alternative approximations for the differences between mean of order r averages, see Vartia (1978; 278-279). Vartia's approximations involve variances of logarithms of prices whereas our approximations involve variances of deflated prices. Our analysis is a variation on his pioneering analysis.

[23] Note that $m_{0,\alpha}(p)$ can be regarded as a weighted Jevons (1865) price level or a Cobb Douglas (1928) price level. Similarly, $p_\alpha \equiv m_{1,\alpha}(p)$ can be regarded as a weighted Dutot (1738) price level or a Leontief (1936) price level.

$\leq \Sigma_{n=1}^{N} \alpha_n p_n$.

Thus the bigger is the variation in the N prices $p_1,...,p_N$, the bigger will be $\mathrm{Var}_\alpha(p/p_\alpha)$ and the more the weighted arithmetic mean of the prices, $\Sigma_{n=1}^{N} \alpha_n p_n$, will be greater than the corresponding weighted geometric mean of the prices, $\prod_{n=1}^{N} (p_n)^{\alpha_n}$. Note that if all of the p_n are equal, then $\mathrm{Var}_\alpha(p/p_\alpha)$ will be equal to 0 and the approximations in (6) and (7) become exact equalities.

Recall that the unit value price and quantity sold for product n during period t was defined as p_{tn} and q_{tn} for n = 1,...,N and t = 1,...,T. At this point, it is useful to define the Jevons (1865) and Dutot (1738) period t *price levels* for the prices in our window of observations, p_J^t and p_D^t, and the corresponding Jevons and Dutot *price indexes*, P_J^t and P_D^t, for t = 1,...,T:

(8) $p_D^t \equiv \Sigma_{n=1}^{N} (1/N)p_{tn}$;
(9) $p_J^t \equiv \Pi_{n=1}^{N} p_{tn}^{1/N}$;
(10) $P_D^t \equiv p_D^t/p_D^1$;
(11) $P_J^t \equiv p_J^t/p_J^1 = \Pi_{n=1}^{N} (p_{tn}/p_{1n})^{1/N}$.

Thus the period t price index is simply the period t price level divided by the corresponding period 1 price level. Note that the Jevons price index can also be written as the geometric mean of the long term price ratios (p_{tn}/p_{1n}) between the period t prices relative to the corresponding period 1 prices.

The *weighted Dutot and Jevons period t price levels* using a weight vector α which satisfies the restrictions (1), $p_{D\alpha}^t$ and $p_{J\alpha}^t$, are defined by (12) and (13) and the corresponding *weighted Dutot and Jevons period t price indexes*, $P_{D\alpha}^t$ and $P_{J\alpha}^t$,[24] are defined by (14) and (15) for t = 1,...,T:

(12) $p_{D\alpha}^t \equiv \Sigma_{n=1}^{N} \alpha_n p_{tn} = m_{1,\alpha}(p^t)$;
(13) $p_{J\alpha}^t \equiv \prod_{n=1}^{N} (p_{tn})^{\alpha_n} = m_{0,\alpha}(p^t)$;
(14) $P_{D\alpha}^t \equiv p_{D\alpha}^t/p_{D\alpha}^1 = \alpha\cdot p^t/\alpha\cdot p^1$;
(15) $P_{J\alpha}^t \equiv p_{J\alpha}^t/p_{J\alpha}^1 = \prod_{n=1}^{N} (p_{tn}/p_{1n})^{\alpha_n}$.

Obviously, (12)-(15) reduce to definitions (8)-(11) if $\alpha = (1/N)1_N$. We can use the approximation (7) for $p = p^1$ and $p = p^t$ in order to obtain the following approximate relationship between the weighted Dutot price index for period t, $P_{D\alpha}^t$, and the corresponding weighted Jevons index, P_J^t:

(16) $P_{J\alpha}^t \equiv p_{J\alpha}^t/p_{J\alpha}^1$;   t = 1,...,T
$= m_{0,\alpha}(p^t)/m_{0,\alpha}(p^1)$ using (2) and (13)
$\approx m_{1,\alpha}(p^t)\{1 - (½)\Sigma_{n=1}^{N} \alpha_n[(p_{tn}/p_\alpha^t) - 1]^2\}/m_{1,\alpha}(p^1)\{1 - (½)\Sigma_{n=1}^{N} \alpha_n[(p_{1n}/p_\alpha^1) - 1]^2\}$
using (7) for $p = p^t$ and $p = p^1$ where $p_\alpha^t \equiv \alpha\cdot p^t$ and $p_\alpha^1 \equiv \alpha\cdot p^1$

[24] This type of index is frequently called a *Geometric Young index*; see Armknecht and Silver (2014; 4-5).

$$= P_{D\alpha}{}^t\{1 - (½)\Sigma_{n=1}{}^N\ \alpha_n[(p_{tn}/p_\alpha{}^t) - 1]^2\}/\{1 - (½)\Sigma_{n=1}{}^N\ \alpha_n[(p_{1n}/p_\alpha{}^1) - 1]^2\}$$
$$= P_{D\alpha}{}^t\{1 - (½)\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)\}/\{1 - (½)\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)\}.$$

In the elementary index context where there are no trends in prices in diverging directions, it is likely that $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)$ is approximately equal to $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$.[25] Under these conditions, *the weighted Jevons price index* $P_{J\alpha}{}^t$ *is likely to be approximately equal to the corresponding weighted Dutot price index*, $P_{D\alpha}{}^t$. Of course, this approximate equality result extends to the case where $\alpha = (1/N)1_N$ and so it is likely that the Dutot price indexes $P_D{}^t$ are approximately equal to their Jevons price index counterparts, $P_J{}^t$.[26] However, if the variance of the deflated period 1 prices is unusually large (small), then there will be a tendency for $P_J{}^t$ to exceed (to be less than) $P_D{}^t$ for $t > 1$.[27]

At higher levels of aggregation where the products may not be very similar, it is likely that there will be *divergent trends in prices* over time. In this case, we can expect $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)$ to exceed $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$. Thus using (16) under these circumstances leads to the likelihood that the weighted index $P_{J\alpha}{}^t$ will be significantly lower than $P_{D\alpha}{}^t$. Similarly, under the *diverging trends in prices hypothesis*, we can expect the ordinary Jevons index $P_J{}^t$ to be lower than the ordinary Dutot index $P_D{}^t$.[28]

We conclude this section by finding an approximate relationship between a CES price index and the corresponding weighted Dutot price index $P_{D\alpha}{}^t$. This approximation result assumes that econometric estimates for the parameters of the CES unit cost function $m_{r,\alpha}(p)$ defined by (2) are available so that we have estimates for the weighting vector α (which we assume satisfies the restrictions (1)) and the parameter r which we assume satisfies $r \leq 1$.[29] The *CES period t price levels* using a weight vector α which satisfies the restrictions (1) and an $r \leq 1$, $p_{CES\alpha}{}^t$, and the corresponding *CES period t price indexes*, $P_{CES\alpha,r}{}^t$, are defined as follows for t = 1,...,T:

---

[25] Note that the vectors $p^t/p_\alpha{}^t$ and $p_\alpha{}^1/p_\alpha{}^1$ are price vectors that are divided by their α weighted arithmetic means. Thus these vectors have eliminated general inflation between periods 1 and t.

[26] The same approximate inequalities hold for the weighted case. An approximation result similar to (16) for the equal weights case where $\alpha = (1/N)1_N$ was first obtained by Carruthers, Sellwood and Ward (1980; 25).

[27] For our empirical example considered in Appendix 1, we found that the sample mean of the $P_D{}^t$ was 0.94581 while the sample mean of the $P_J{}^t$ was 0.94956 so that in general, the Dutot price index was *below* the corresponding Jevons index. However, the period 1 variance, $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$ (for $\alpha = (1/N)1_N$) was unusually large (equal to 0.072) as compared to the sample average of the $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)$ (equal to 0.063) and this explains why $P_D{}^t$ was generally below $P_J{}^t$ for our particular data set. In general, for highly substitutable products, we expect the Jevons price index to lie below its Dutot counterpart. When we dropped the data for the first year, we found that the resulting sample mean of the $P_D{}^t$ was 0.86080 while the sample mean of the $P_J{}^t$ was 0.86664 so that in this case, the Dutot price index was *above* the corresponding Jevons index (as expected). The data for the first year in our sample was unusual due to the absence of products 2 and 4 for most of the year.

[28] Furthermore, as we shall see later, the Dutot index can be viewed as a fixed basket index where the basket is a vector of ones. Thus it is subject to substitution bias which will show up under the divergent price trends hypothesis.

[29] These restrictions imply that $m_{r,\alpha}(p)$ is a linearly homogeneous, nondecreasing and concave function of the price vector p. These restrictions must be satisfied if we apply the economic approach to price index theory.

(17) $p_{CES\alpha,r}{}^t \equiv [\Sigma_{n=1}{}^N \alpha_n p_{tn}{}^r]^{1/r} = m_{r,\alpha}(p^t)$;
(18) $P_{CES\alpha,r}{}^t \equiv p_{CES\alpha,r}{}^t/p_{CES\alpha,r}{}^1 = m_{r,\alpha}(p^t)/m_{r,\alpha}(p^1)$.

Now use the approximation (6) for $p = p^1$ and $p = p^t$ in order to obtain the following approximate relationship between the weighted Dutot price index for period t, $P_{D\alpha}{}^t$, and the corresponding period t CES index, $P_{CES\alpha,r}{}^t$ for t = 1,...,T:

$$\begin{aligned}
(19)\ P_{CES\alpha,r}{}^t &\equiv p_{CES\alpha,r}{}^t/p_{CES\alpha,r}{}^1 ; \\
&= m_{r,\alpha}(p^t)/m_{r,\alpha}(p^1) && \text{using (17) and (18)} \\
&\approx [m_{1,\alpha}(p^t)/m_{1,\alpha}(p^1)][1 + (½)(r-1)\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)]/[1 + (½)(r-1)\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)] \\
&= P_{D\alpha}{}^t\{1 + (½)(r-1)\Sigma_{n=1}{}^N \alpha_n[(p_{tn}/p_\alpha{}^t) - 1]^2\}/\{1 + (½)(r-1)\Sigma_{n=1}{}^N \alpha_n[(p_{1n}/p_\alpha{}^1) - 1]^2\}
\end{aligned}$$

where we used definitions (4), (12) and (14) to establish the last equality in (19). Again, in the elementary index context with no diverging trends in prices, we could expect $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t) \approx \mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$ for t = 2,...,T. Using this assumption about the approximate constancy of the (weighted) variance of the deflated prices over time, and using (16) and (19), we obtain the following approximations for t = 2,3,...,T:

(20) $P_{CES\alpha,r}{}^t \approx P_{J\alpha}{}^t \approx P_{D\alpha}{}^t$.

Thus under the assumption of approximately *constant variances* for deflated prices, the CES, weighted Jevons and weighted Dutot price indexes should approximate each other fairly closely, provided that the same weighting vector α is used in the construction of these indexes.[30]

The parameter r which appears in the definition of the CES unit cost function is related to the *elasticity of substitution* σ; i.e., it turns out that $\sigma = 1 - r$.[31] Thus as r takes on values from 1 to $-\infty$, σ will take on values from 0 to $+\infty$. In the case where the products are closely related, typical estimates for σ range from 1 to 10 when CES preferences are estimated. Thus if we substitute $\sigma = 1 - r$ into the approximation (19), we obtain the following approximations for t = 1,...,T:

(21) $P_{CES\alpha,r}{}^t \approx P_{D\alpha}{}^t\, [1 - (½)\sigma\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)]/[1 - (½)\sigma\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)]$.

The approximations in (21) break down for large and positive σ (or equivalently, for very negative r); i.e., the expressions in square brackets on the right hand sides of (21) will pass through 0 and become meaningless as σ becomes very large. The approximations become increasingly accurate as σ approaches 0 (or as r approaches 1). Of course, the approximations also become more accurate as the dispersion of prices within a period becomes smaller. For σ between 0 and 1 and with "normal" dispersion of prices, the

[30] Again, the approximate relationship $P_{CES\alpha,r}{}^t \approx P_{D\alpha}{}^t$ may not hold if the variance of the prices in the base period, $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$, is unusually large or small.
[31] See for example, Feenstra (1994; 158) or Diewert and Feenstra (2017).

approximations in (21) should be reasonably good. However, as σ becomes larger, the expressions in square brackets will become closer to 0 and the approximations in (21) will become more volatile and less accurate as σ increases from an initial 0 value.

At higher levels of aggregation where the products are not similar, it is likely that there will be *divergent trends in prices* over time and in this case, we can expect $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)$ to exceed $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$. In this case, the approximate *equalities* (20) will no longer hold. In the case where the elasticity of substitution σ is greater than 1 (so $r < 0$) and $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t) > \mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$, we can expect that $P_{CES\alpha,r}{}^t < P_{D\alpha}{}^t$ and the gaps between these two indexes will grow bigger over time as $\mathrm{Var}_\alpha(p^t/p_\alpha{}^t)$ grows larger than $\mathrm{Var}_\alpha(p^1/p_\alpha{}^1)$.

In the following section, we will use the mean of order r function to aggregate the price ratios p_{tn}/p_{1n} into an aggregate price index for period t directly; i.e., we will not construct *price levels* as a preliminary step in the construction of a *price index*.

3. Using Means of Order r to Aggregate Price Ratios

In the previous section, we compared various elementary indexes using approximate relationships between price levels constructed by using means of order r to aggregate prices. In this section, we will develop approximate relationships between price indexes constructed by using means of order r defined over price ratios.

In what follows, it is assumed that the weight vector α satisfies conditions (1); i.e., $\alpha >> 0_N$ and $\alpha\cdot 1_N = 1$. Define the *mean of order r price index for period t* (relative to period 1), $P_{r,\alpha}{}^t$, as follows for t = 1,...,T:

$$(22)\quad P_{r,\alpha}{}^t \equiv [\Sigma_{n=1}{}^N \alpha_n(p_{tn}/p_{1n})^r]^{1/r}\ ;\ r \neq 0;$$
$$\equiv \prod_{n=1}{}^N (p_{tn}/p_{1n})^{\alpha_n}\qquad ;\ r = 0.$$

When $r = 1$ and $\alpha = (1/N)1_N$, then $P_{r,\alpha}{}^t$ becomes the *period t fixed base Carli* (1804) *price index*, $P_C{}^t$, defined as follows for t = 1,...,T:

$$(23)\quad P_C{}^t \equiv \Sigma_{n=1}{}^N (1/N)(p_{tn}/p_{1n}).$$

With a general α and $r = 1$, $P_{r,\alpha}{}^t$ becomes the *fixed base weighted Carli price index*, $P_{C\alpha}{}^t$,[32] defined as follows for t = 1,...,T:

$$(24)\quad P_{C\alpha}{}^t \equiv \Sigma_{n=1}{}^N \alpha_n(p_{tn}/p_{1n}).$$

Using (24), it can be seen that the α weighted mean of the period t long term price ratios p_{tn}/p_{1n} divided by $P_{C\alpha}{}^t$ is equal to 1; i.e., we have for t = 1,...,T:

[32] This type of index is due to Arthur Young (1812; 72) and so we could call this index the *Young index*, $P_{Y\alpha}{}^t$.

(25) $\Sigma_{n=1}^{N} \alpha_n(p_{tn}/p_{1n}P_{C\alpha}^{t}) = 1$.

Denote the α weighted variance of the deflated period t price ratios $p_{tn}/p_{1n}P_{C\alpha}^{t}$ as $\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t})$ and define it as follows for t = 1,...,T:

(26) $\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t}) \equiv \Sigma_{n=1}^{N} \alpha_n[(p_{tn}/p_{1n}P_{C\alpha}^{t}) - 1]^2$.

Proposition 2: Let $p >> 0_N$, $\alpha >> 0_N$ and $\alpha\cdot 1_N = 1$. Then $P_{r,\alpha}^{t}/P_{1,\alpha}^{t} = P_{r,\alpha}^{t}/P_{C\alpha}^{t}$ is approximately equal to the following expression for any r for t = 1,...,T:

(27) $P_{r,\alpha}^{t}/P_{C\alpha}^{t} \approx 1 + (½)(r - 1)\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t})$

where $P_{r,\alpha}^{t}$ is the mean of order r price index (with weights α) defined by (22), $P_{C\alpha}^{t}$ is the α weighted Carli index defined by (24) and $\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t})$ is the α weighted variance of the deflated long term price ratios $(p_{tn}/p_{1n})/P_{C\alpha}^{t}$ defined by (26).

Proof: Replace the vector p in Proposition 1 by the vector $[p_{t1}/p_{11},p_{t2}/p_{12},...,p_{tN}/p_{1N}]$.[33] Then the ratio $m_{r,\alpha}(p)/m_{1,\alpha}(p)$ which appears on the left hand side of (5) becomes the ratio $P_{r,\alpha}^{t}/P_{1,\alpha}^{t} = P_{r,\alpha}^{t}/P_{C\alpha}^{t}$ using definitions (22) and (24). The terms p_α and $\mathrm{Var}_\alpha(p/p_\alpha)$ which appear on the right hand side of (5) become $P_{C\alpha}^{t}$ and $\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t})$ respectively. With these substitutions, (5) becomes (27) and we have established Proposition 2. Q.E.D.

It is useful to look at the special case of (27) when r = 0. In this case, using definitions (22) and (15), we can establish the following equalities for t = 1,...,T:

(28) $P_{0,\alpha}^{t} \equiv \prod_{n=1}^{N} (p_{tn}/p_{1n})^{\alpha_n} = P_{J\alpha}^{t}$

where $P_{J\alpha}^{t}$ is the period t *weighted Jevons* or *Cobb Douglas price index* defined by (15) in the previous section.[34] Thus when r = 0, the approximations defined by (27) become the following approximations for t = 1,...,T:

(29) $P_{J\alpha}^{t}/P_{C\alpha}^{t} \approx 1 - (½)\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^{t})$.

Thus the bigger is the α weighted variance of the deflated period t long term price ratios, $(p_{t1}/p_{11})/P_{C\alpha}^{t}$,..., $(p_{tN}/p_{1N})/P_{C\alpha}^{t}$, the more the period t weighted Carli index $P_{C\alpha}^{t}$ will exceed the corresponding period t weighted Jevons index $P_{J\alpha}^{t}$.

When $\alpha = (1/N)1_N$, the approximations (29) become the following approximate relationships between the period t *Carli index* P_C^{t} defined by (23) and the period t *Jevons index* P_J^{t} defined by (11) for t = 1,...,T:[35]

[33] In Proposition 1, some prices in either period could be 0. However, Proposition 2 requires that all period 1 prices be positive.
[34] Again, recall that Armknecht and Silver (2014; 4) call this index the Geometric Young index.
[35] Results that are essentially equivalent to (30) were first obtained by Dalén (1992) and Diewert (1995). The approximations in (27) and (29) for weighted indexes are new. Vartia and Suoperä (2018; 5) derived

(30) $P_J^t/P_C^t \approx 1 - (½)\mathrm{Var}_{(1/N)1}(p^t/p^1P_C^t)$
$= 1 - (½)\Sigma_{n=1}^N (1/N)[(p_{tn}/p_{1n}P_C^t) - 1]^2$.

Thus the Carli price indexes P_C^t will exceed their Jevons counterparts P_J^t (unless $p^t = \lambda_t p^1$ in which case prices in period t are proportional to prices in period 1 and in this case, $P_C^t = P_J^t$). This is an important result since from an axiomatic perspective, the Jevons price index has much better properties than the corresponding Carli indexes[36] and in particular, *chaining Carli indexes will lead to large upward biases as compared to their Jevons counterparts*.[37]

The results in this section can be summarized as follows: holding the weight vector α constant, the weighted Jevons price index for period t, $P_{J\alpha}^t$ is likely to lie below the corresponding weighted Carli index, $P_{C\alpha}^t$, with the gap growing as the α weighted variance of the deflated price ratios, $(p_{t1}/p_{11})/P_{C\alpha}^t,..., (p_{tN}/p_{1N})/P_{C\alpha}^t$, increases.[38]

In the following section, we turn our attention to weighted price indexes where the weights are not exogenous constants but depend on observed sales shares.

4. Relationships between Some Share Weighted Price Indexes

In this section (and in subsequent sections), we will look at comparisons between price indexes that use information on the observed expenditure or sales shares of products in addition to price information. Recall that $s_{tn} \equiv p_{tn}q_{tn}/p^t\cdot q^t$ for n = 1,...,N and t = 1,...,T.

The *fixed base Laspeyres* (1871) *price index* for period t, P_L^t, is defined as the following base period share weighted *arithmetic* average of the price ratios, p_{tn}/p_{1n}, for t = 1,...,T:

(31) $P_L^t \equiv \Sigma_{n=1}^N s_{1n}(p_{tn}/p_{1n})$.

It can be seen that P_L^t is a weighted Carli index $P_{C\alpha}^t$ of the type defined by (34) in the previous section where $\alpha \equiv s^1 \equiv [s_{11},s_{12},...,s_{1N}]$. We will compare P_L^t with its weighted geometric mean counterpart, P_{GL}^t, which is a weighted Jevons index $P_{J\alpha}^t$ where the

---

alternative approximations. The analysis in this section is similar to Vartia's (1978; 276-289) analysis of Fisher's (1922) five-tined fork.

[36] See Diewert (1995) on this topic.

[37] The same upward bias holds for weighted Carli indexes relative to their weighted Jevons counterparts. For our frozen juice data listed in Appendix A, the 3 year sample average of the Carli indexes was 0.96277 as compared to the corresponding sample average of the Jevons indexes which was 0.94956.

[38] Since the Jevons price index has the best axiomatic properties, this result implies that CPI compilers should avoid the use of the Carli index in the construction of a CPI. This advice goes back to Fisher (1922; 29-30). Since the Dutot index will approximate the corresponding Jevons index provided that the products are similar and there are no systematic divergent trends in prices, Dutot indexes can be satisfactory at the elementary level. If the products are not closely related, Dutot indexes become problematic since they are not invariant to changes in the units of measurement. Moreover, in the case of nonsimilar products, divergent trends in prices become more probable and thus the Dutot index will tend to be above the corresponding Jevons index due to substitution bias.

weight vector is $\alpha = s^1$. Thus the logarithm of the *fixed base Geometric Laspeyres price index* is defined as follows for t = 1,...,T:[39]

(32) $\ln P_{GL}^t \equiv \Sigma_{n=1}^N s_{1n} \ln(p_{tn}/p_{1n})$.

Since P_{GL}^t and P_L^t are weighted geometric and arithmetic means of the price ratios p_{tn}/p_{1n} (using the weights in the period 1 share vector s^1), Schlömilch's inequality implies that $P_{GL}^t \leq P_L^t$ for t = 1,...,T. The inequalities (29), with $\alpha = s^1$, give us approximations to the gaps between the $P_{GL}^t = P_{J\alpha}^t$ and the $P_{C\alpha}^t = P_L^t$. Thus we have the following approximate equalities for $\alpha = s^1$ and t = 1,...,T:

(33) $P_{GL}^t/P_L^t \approx 1 - (½)\mathrm{Var}_\alpha(p^t/p^1P_L^t) = 1 - (½)\Sigma_{n=1}^N s_{1n}[(p_{tn}/p_{1n}P_L^t) - 1]^2$.

Using our scanner data set on 19 frozen fruit juice products listed in the Appendix for 39 months, we found that $\mathrm{Var}_\alpha(p^t/p^1P_{C\alpha}^t)$ for $t \geq 2$ varied between 0.0024 and 0.0627 with a mean of 0.0234. Thus there was a considerable amount of variation in these variance terms. The sample mean of the ratios P_{GL}^t/P_L^t was 0.9873 so that P_{GL}^t was below P_L^t by 1.27 percentage points on average. The sample mean of the error terms on the right hand sides of the approximations defined by (33) was 0.9883 which is only below the sample mean of the ratios P_{GL}^t/P_L^t by 0.1 percentage points. The correlation coefficient between the ratios P_{GL}^t/P_L^t and the corresponding error terms on the right hand sides of (33) was 0.9950. Thus the approximate equalities in (33) were quite close to being equalities.

The *fixed base Paasche* (1874) *price index* for period t, P_P^t, is defined as the following period t share weighted *harmonic* average of the price ratios, p_{tn}/p_{1n}, for t = 1,...,T:

(34) $P_P^t \equiv [\Sigma_{n=1}^N s_{tn}(p_{tn}/p_{1n})^{-1}]^{-1}$.

We will compare P_P^t with its weighted geometric mean counterpart, P_{GP}^t, which is a weighted Jevons index $P_{J\alpha}^t$ where the weight vector is $\alpha = s^t$. Thus the logarithm of the *fixed base Geometric Paasche price index* is defined as follows for t = 1,...,T:

(35) $\ln P_{GP}^t \equiv \Sigma_{n=1}^N s_{tn} \ln(p_{tn}/p_{1n})$.

Since P_{GP}^t and P_P^t are weighted geometric and harmonic means of the price ratios p_{tn}/p_{1n} (using the weights in the period t share vector s^t), Schlömilch's inequality implies that $P_P^t \leq P_{GP}^t$ for t = 1,...,T. However, we cannot apply the inequalities (29) directly to give us an approximation to the *size* of the gap between P_{GP}^t and P_P^t. Viewing definition (34), it can be seen that the reciprocal of P_P^t is a period t share weighted average of the reciprocals of the long term price ratios, p_{11}/p_{t1}, p_{12}/p_{t2}, ..., p_{1N}/p_{tN}. Thus using definition (34), we have the following equations and inequalities for $\alpha = s^t$ and t = 1,...,T:

(36) $[P_P^t]^{-1} = \Sigma_{n=1}^N s_{tn}(p_{1n}/p_{tn})$

[39] Vartia (1978; 272) used the terms "geometric Laspeyres" and "geometric Paasche" to describe the indexes defined by (32) and (35).

$\geq \prod_{n=1}^{N} (p_{1n}/p_{tn})^{s_{tn}}$
$= [P_{GP}^{t}]^{-1}$ using definitions (35)

where the inequalities in (36) follow from Schlömilch's inequality; i.e., a weighted arithmetic mean is always equal to or greater than the corresponding weighted geometric mean. Note that the first equation in (36) implies that the period t share weighted mean of the reciprocal price ratios, p_{1n}/p_{tn}, is equal to the reciprocal of P_P^t. Now adapt the approximate equalities (29) in order to establish the following approximate equalities for t = 1,...,T:

(37) $[P_{GP}^{t}]^{-1}/[P_P^t]^{-1} \approx 1 - (½)\Sigma_{n=1}^{N} s_{tn}[(p_{1n}/p_{tn}\,[P_P^t]^{-1}) - 1]^2$.

The approximate equalities (37) may be rewritten as follows for t = 1,...,T:

(38) $P_{GP}^{t} \approx P_P^t/\{1 - (½)\Sigma_{n=1}^{N} s_{tn}[(p_{1n}P_P^t/p_{tn}) - 1]^2\}$.

Thus for t = 1,...,T, we have $P_{GP}^{t} \geq P_P^t$ (and the approximate equalities (38) measure the gaps between these indexes) and $P_{GL}^{t} \leq P_L^t$ (and the approximate equalities (33) measure the gaps between these indexes). Later we will show that the inequalities $P_{GP}^{t} \leq P_{GL}^{t}$ are likely if the N products are close substitutes for each other.

Using our scanner data set listed in the Appendix, we found that the variance terms on the right hand sides of (38), $\Sigma_{n=1}^{N} s_{tn}[(p_{1n}P_P^t/p_{tn}) - 1]^2$, for $t \geq 2$ varied between 0.0026 and 0.0688 with a mean of 0.0330. Thus these variances were bigger than the corresponding variances in (33) by one percentage point on average. This means that on average, $P_{GL}^{t}/P_L^t > P_P^t/P_{GP}^{t}$ or on average for our sample, $P_{GL}^{t}P_{GP}^{t} > P_L^t P_P^t$, which in turn implies that on average, $P_T^t > P_F^t$ for our sample.[40] The sample mean of the ratios P_{GP}^{t}/P_P^t was 1.0169 so that P_{GP}^{t} was above P_P^t by 1.69 percentage points on average. The sample mean of the error terms $1/\{1 - (½)\Sigma_{n=1}^{N} s_{tn}[(p_{1n}P_P^t/p_{tn}) - 1]^2\}$ on the right hand sides of the approximations defined by (38) was 1.0169 which is identical to the sample mean of the ratios P_{GP}^{t}/P_P^t. The correlation coefficient between the ratios P_{GP}^{t}/P_P^t and the corresponding error terms on the right hand sides of (38) was 0.9922. Thus the approximate equalities in (38) were quite close to being equalities.

Suppose that prices in period t are proportional to the corresponding prices in period 1 so that $p^t = \lambda_t p^1$ where λ_t is a positive scalar. Then it is straightforward to show that $P_P^t = P_{GP}^{t} = P_{GL}^{t} = P_L^t = \lambda_t$ and the error terms for equation t in (34) and (39) are equal to 0.

Define the period t fixed base Fisher (1922) and Törnqvist[41] price indexes, P_F^t and P_T^t, as the following geometric means for t = 1,...,T:

(39) $P_F^t \equiv [P_L^t P_P^t]^{1/2}$;
(40) $P_T^t \equiv [P_{GL}^{t} P_{GP}^{t}]^{1/2}$.

[40] P_F^t and P_T^t are defined by (39) and (40) below.
[41] See Törnqvist (1936) and Törnqvist and Törnqvist (1937) and Theil (1967; 136-137).

Thus P_F^t is the geometric mean of the period t fixed base Laspeyres and Paasche price indexes while P_T^t is the geometric mean of the period t fixed base geometric Laspeyres and geometric Paasche price indexes. Now use the approximate equalities in (34) and (38) and substitute these equalities into (40) in order to obtain the following approximate equalities between P_T^t and P_F^t for ft = 1,...,T:

$$(41)\ P_T^t \equiv [P_{GL}^t P_{GP}^t]^{1/2} \approx [P_L^t P_P^t]^{1/2}\, \varepsilon(p^1,p^t,s^1,s^t) = P_F^t\, \varepsilon(p^1,p^t,s^1,s^t)$$

where the approximation error function $\varepsilon(p^1,p^t,s^1,s^t)$ is defined as follows for t = 1,...,T:

$$(42)\ \varepsilon(p^1,p^t,s^1,s^t) \equiv \{1 - (½)\Sigma_{n=1}^N s_{1n}[(p_{tn}/p_{1n}P_L^t) - 1]^2\}^{1/2}/\{1 - (½)\Sigma_{n=1}^N s_{tn}[(p_{1n}P_P^t/p_{tn}) - 1]^2\}^{1/2}.$$

Thus P_T^t is approximately equal to P_F^t for t = 1,...,T. But how good are these approximations? We know from Diewert (1978) that $P_T^t = P_T(p^1,p^t,s^1,s^t)$ approximates $P_F^t = P_F(p^1,p^t,s^1,s^t)$ to the second order around any point where $p^t = p^1$ and $s^t = s^1$.[42] We also know that the approximations in (33) and (38) are fairly good, at least for our scanner data set. Thus it is likely that the error terms $\varepsilon(p^1,p^t,s^1,s^t)$ are close to 1.[43]

Using our scanner data set listed in the Appendix, we found that the sample mean of the ratios P_T^t/P_F^t was 1.0019 so that P_T^t was above P_F^t by 0.19 percentage points on average. The sample mean of the error terms $\varepsilon(p^1,p^t,s^1,s^t)$ defined by (42) was 1.0025. The correlation coefficient between the ratios P_T^t/P_F^t and the corresponding error terms $\varepsilon(p^1,p^t,s^1,s^t)$ on the right hand sides of (41) was 0.9891. Thus the approximate equalities in (41) were quite close to being equalities. However, if the products were highly substitutable and if prices and shares trended in opposite directions, then we expect that the base period share weighted variance $\Sigma_{n=1}^N s_{1n}[(p_{tn}/p_{1n}P_L^t) - 1]^2$ will increase as t increases and we expect the period t share weighted variance $\Sigma_{n=1}^N s_{tn}[(p_{1n}P_P^t/p_{tn}) - 1]^2$ to increase even more as t increases because as p_{tn} becomes smaller, $[(p_{1n}P_P^t/p_{tn}) - 1]^2$ becomes bigger and the share weight s_{tn} will also increase. Thus P_T will tend to increase relative to P_F as time increases under these conditions. The more substitutable the products are, the greater will be this tendency.

Our tentative conclusion at this point is that the approximations defined by (33), (38) and (41) are good enough to provide rough estimates of the differences in the six price indexes involved in these approximate equalities. Empirically, we found that the variance

---

[42] This result can be generalized to the case where $p^t = \lambda p^1$ and $s^t = s^1$.

[43] However, the Diewert (1978) second order approximation is different from the present second order approximations that are derived from Proposition 2. Thus the closeness of $\varepsilon(p^1,p^t,s^1,s^t)$ to 1 depends on the closeness of the Diewert second order approximation of P_T^t to P_F^t and the closeness of the second order approximations that were used in (33) and (38), which use different Taylor series approximations. Vartia and Suoperä (2018) used alternative Taylor series approximations to obtain relationships between various indexes.

terms on the right hand sides of (38) tended to be larger than the corresponding variances on the right hand sides of (33) and these differences led to a tendency for the fixed base Fisher price indexes P_F^t to be slightly smaller than the corresponding fixed base Törnqvist Theil price indexes P_T^t.[44]

We conclude this section by developing an exact relationship between the geometric Laspeyres and Paasche price indexes. Using definitions (32) and (35) for the logarithms of these indexes, we have the following exact decomposition for the logarithmic difference between these indexes for t = 1,...,T:[45]

$$(43)\ \ln P_{GP}^t - \ln P_{GL}^t = \Sigma_{n=1}^N s_{tn} \ln(p_{tn}/p_{1n}) - \Sigma_{n=1}^N s_{1n} \ln(p_{tn}/p_{1n})$$
$$= \Sigma_{n=1}^N [s_{tn} - s_{1n}][\ln p_{tn} - \ln p_{1n}].$$

Define the vectors $\ln p^t \equiv [\ln p_{t1},\ln p_{t2},...,\ln p_{tN}]$ for t = 1,...,T. It can be seen that the right hand side of equation t equations (43) is equal to $[s^t - s^1]\cdot[\ln p^t - \ln p^1]$, the inner product of the vectors $x \equiv s^t - s^1$ and $y \equiv \ln p^t - \ln p^1$. Let x^* and y^* denote the arithmetic means of the components of the vectors x and y. Note that $x^* \equiv (1/N)1_N\cdot x = (1/N)1_N\cdot[s^t - s^1] = (1/N)[1 - 1] = 0$. The covariance between x and y is defined as $\mathrm{Cov}(x,y) \equiv (1/N)[x - x^*1_N]\cdot[y - y^*1_N] = (1/N)\, x\cdot y - x^*y^* = (1/N)\, x\cdot y$[46] since x^* is equal to 0. Thus the right hand side of (43) is equal to $N\,\mathrm{Cov}(x,y) = N\,\mathrm{Cov}(s^t - s^1,\ln p^t - \ln p^1)$; i.e., the right hand side of (43) is equal to N times the covariance of the long term share difference vector, $s^t - s^1$, with the long term log price difference vector, $\ln p^t - \ln p^1$. Hence if this covariance is positive, then $\ln P_{GP}^t - \ln P_{GL}^t > 0$ and $P_{GP}^t > P_{GL}^t$. If this covariance is negative, then $P_{GP}^t < P_{GL}^t$. We argue below that for the case where the N products are close substitutes, it is likely that the covariances on the right hand side of equations (43) are negative for $t > 1$.

Suppose that the observed price and quantity data are approximately consistent with purchasers having identical Constant Elasticity of Substitution preferences. CES preferences are dual to the CES unit cost function $m_{r,\alpha}(p)$ which is defined by (2) above where α satisfies (1) and $r \leq 1$. It can be shown[47] that the sales share for product n in a period where purchasers face the strictly positive price vector $p \equiv [p_1,...,p_N]$ is the following share:

$$(44)\ s_n(p) \equiv \alpha_n p_n^r/\Sigma_{i=1}^N \alpha_i p_i^r\ ; \quad n = 1,...,N.$$

Upon differentiating $s_n(p)$ with respect to p_n, we find that the following relations hold:

$$(45)\ \partial \ln s_n(p)/\partial \ln p_n = r[1-s_n(p)]\ ; \quad n = 1,...,N.$$

---

[44] Vartia and Suoperä (2018) also found a tendency for the Fisher price index to lie slightly below their Törnqvist counterparts in their empirical work.
[45] Vartia and Suoperä (2018; 26) derived this result and noticed that the right hand side of (43) could be interpreted as a covariance. They also developed several alternative exact decompositions for the difference $\ln P_{GP}^t - \ln P_{GL}^t$. Their paper also develops a new theory of "excellent" index numbers.
[46] This equation is the *covariance identity* which was first used by Bortkiewicz (1923) to show that normally the Paasche price index is less than the corresponding Laspeyres index.
[47] See Diewert and Feenstra (2017) for example.

Thus $\partial \ln s_n(p)/\partial \ln p_n < 0$ if $r < 0$ (or equivalently, if the elasticity of substitution $\sigma \equiv 1 - r$ is greater than 1) and $\partial \ln s_n(p)/\partial \ln p_n > 0$ if r satisfies $0 < r < 1$ (or equivalently, if the elasticity of substitution satisfies $0 < \sigma < 1$). If we are aggregating prices at the first stage of aggregation where the products are close substitutes and purchasers have common CES preferences, then it is likely that the elasticity of substitution is greater than 1 and hence as the price of product n decreases, it is likely that the share of that product will increase. Hence we expect the terms $[s_{tn} - s_{1n}][\ln p_{tn} - \ln p_{1n}]$ to be predominantly negative; i.e., if p_{1n} is unusually low, then $\ln p_{tn} - \ln p_{1n}$ is likely to be positive and $s_{tn} - s_{1n}$ is likely to be negative. On the other hand, if p_{tn} is unusually low, then $\ln p_{tn} - \ln p_{1n}$ is likely to be negative and $s_{tn} - s_{1n}$ is likely to be positive. Thus for closely related products, we expect the covariances on the right hand sides of (43) to be negative and for P_{GP}^t to be less than P_{GL}^t. We can combine this inequality with our previously established inequalities to conclude that for closely related products, it is likely that $P_P^t < P_{GP}^t < P_T^t < P_{GL}^t < P_L^t$. On the other hand, if we are aggregating at higher levels of aggregation, then it is likely that the elasticity of substitution is in the range $0 < \sigma < 1$,[48] and in this case, the covariances on the right hand sides of (43) will tend to be positive and hence in this case, it is likely that $P_{GP}^t > P_{GL}^t$. We also have the inequalities $P_P^t < P_{GP}^t$ and $P_{GL}^t < P_L^t$ in this case.[49]

We turn now to some relationships between weighted and unweighted (i.e., equally weighted) geometric price indexes.

5. Relationships between the Jevons, Geometric Laspeyres, Geometric Paasche and Törnqvist Price Indexes

In this section, we will investigate how close the unweighted Jevons index P_J^t is to the geometric Laspeyres P_{GL}^t, geometric Paasche P_{GP}^t and Törnqvist P_T^t price indexes.

We first investigate the difference between the logarithms of P_{GL}^t and P_J^t. Using the definitions for these indexes, we have the following log differences for t = 1,...,T:

$$(46)\quad \begin{aligned} \ln P_{GL}^t - \ln P_J^t &= \Sigma_{n=1}^N [s_{1n} - (1/N)][\ln p_{tn} - \ln p_{1n}] \\ &= N\mathrm{Cov}(s^1 - (1/N)1_N, \ln p^t - \ln p^1) \\ &\equiv \varepsilon_t. \end{aligned}$$

In the elementary index context where the N products are close substitutes and product shares in period 1 are close to being equal, it is likely that ε_t is positive; i.e., if $\ln p_{1n}$ is unusually low, then s_{1n} is likely to be unusually high and thus it is likely that $s_{1n} - (1/N) >$

[48] See Shapiro and Wilcox (1997) who found that $\sigma = 0.7$ fit the US data well at higher levels of aggregation. See also Armknecht and Silver (2014; 9) who noted that estimates for σ tend to be greater than 1 at the lowest level of aggregation and less than 1 at higher levels of aggregation.

[49] See Vartia (1978; 276-290) for a similar discussion about the relationships between P_L^t, P_P^t, P_F^t, P_{GL}^t, P_{GP}^t and P_T^t. Vartia extended the discussion to include period 1 and period t share weighted harmonic averages of the price ratios, p_{tn}/p_{1n}. See also Armknecht and Silver (2014; 10) for a discussion on how weighted averages of the above indexes could approximate a superlative index at higher levels of aggregation.

0 and $\ln p_{tn} - \ln p_{1n}$ minus the mean of the log ratios $\ln(p_{tn}/p_{1n})$ is likely to be greater than 0 and hence ε_t is likely to be greater than 0, implying that $P_{GL}^t > P_J^t$. However, if N is small and the shares have a high variance, then if product n goes on sale in period 1, we cannot assert that s_{1n} is likely to be greater than 1/N and hence we cannot be confident that ε_t is likely to be greater than 0 and hence we cannot confidently predict that P_{GL}^t will be greater than P_J^t.

There are three simple sets of conditions that will imply that $P_{GL}^t = P_J^t$: (i) the covariance on the right hand side of (46) equals 0; i.e., $\mathrm{Cov}(s^1 - (1/N)1_N, \ln p^t - \ln p^1) = 0$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) equal sales shares in period 1; i.e., $s^1 = (1/N)1_N$.

Now look at the difference between the logarithms of P_{GP}^t and P_J^t. Using the definitions for these indexes, for t = 1,...,T, we have:

$$
\begin{aligned}
(47)\ \ln P_{GP}^t - \ln P_J^t &= \Sigma_{n=1}^N [s_{tn} - (1/N)][\ln p_{tn} - \ln p_{1n}] \\
&= N\mathrm{Cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) \\
&\equiv \eta_t .
\end{aligned}
$$

In the elementary index context where the N products are close substitutes and the shares s^t are close to being equal, then it is likely that η_t is negative; i.e., if $\ln p_{tn}$ is unusually low, then s_{tn} is likely to be unusually high and thus it is likely that $s_{tn} - (1/N) > 0$ and $\ln p_{tn} - \ln p_{1n}$ minus the mean of the log ratios $\ln(p_{tn}/p_{1n})$ is likely to be less than 0 and hence η_t is likely to be less than 0 implying that $P_{GP}^t < P_J^t$. However if N is small and the period t shares s^t are not close to being equal, then again, we cannot confidently predict the sign of the covariance in (47).

There are three simple sets of conditions that will imply that $P_{GP}^t = P_J^t$: (i) the covariance on the right hand side of (47) equals 0; i.e., $\mathrm{Cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) = 0$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) equal sales shares in period t; i.e., $s^t = (1/N)1_N$.

Using the definitions for P_T^t and P_J^t, the log difference between these indexes is equal to the following expression for t = 1,...,T:

$$
\begin{aligned}
(48)\ \ln P_T^t - \ln P_J^t &= \Sigma_{n=1}^N [(½)s_{tn} + (½)s_{1n} - (1/N)][\ln p_{tn} - \ln p_{1n}] \\
&= N\mathrm{Cov}[(½)s^t + (½)s^1 - (1/N)1_N, \ln p^t - \ln p^1] \\
&= (N/2)\mathrm{Cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) + (N/2)\mathrm{Cov}(s^1 - (1/N)1_N, \ln p^t - \ln p^1) \\
&= (½)\varepsilon_t + (½)\eta_t .
\end{aligned}
$$

As usual, there are three simple sets of conditions that will imply that $P_T^t = P_J^t$: (i) the covariance on the right hand side of (48) equals 0; i.e., $\mathrm{Cov}[(½)s^t + (½)s^1 - (1/N)1_N, \ln p^t - \ln p^1] = 0 = (½)\varepsilon_t + (½)\eta_t$ or equivalently, $\mathrm{Cov}(s^t - (1/N)1_N, \ln p^t - \ln p^1) = -\mathrm{Cov}(s^t - (1/N)1_N, \ln p^1 - \ln p^1)$; (ii) period t price proportionality; i.e., $p^t = \lambda_t p^1$ for some $\lambda_t > 0$; (iii) the arithmetic average of the period 1 and t sales shares are all equal to 1/N; i.e., $(½)s^t + (½)s^1 = (1/N)1_N$.

If the trend deflated prices p_{tn}/λ_t are distributed *independently* across time and *independently* of the sales shares s_{tn}, then it can be seen that the expected values of the ε_t and η_t will be 0 and hence $P_T^t \approx P_J^t$ for t = 1,...,T. Thus it can be the case that the ordinary Jevons price index is able to provide an adequate approximation to the superlative Törnqvist price index in the elementary price index context. However, if the shares are trending and if prices are trending in divergent directions, then P_J^t will not be able to approximate P_T^t. To illustrate this point, assume that shares and the logarithms of prices are *trending linearly over time*; i.e., make the following assumptions:

(49) $s^t = s^1 + \beta(t-1)$; $\ln p^t = \ln p^1 + \gamma(t-1)$; t = 2,3,...,T

where $\beta \equiv [\beta_1,...,\beta_N] \equiv [\beta_1,...,\beta_N]$ and $\gamma \equiv [\gamma_1,...,\gamma_N]$ are constant vectors and β satisfies the additional restriction:

(50) $\beta \cdot 1_N = 0$.

The restriction (50) is required to ensure that the shares s^t sum to unity.[50] Substitute assumptions (49) into equations (46) and we obtain the following equations for t = 2,3,...,T:

$$(51)\ \ln P_{GL}^t - \ln P_J^t = (s^1 - (1/N)1_N)\cdot(\ln p^t - \ln p^1)$$
$$= (s^1 - (1/N)1_N)\cdot\gamma(t-1).$$

Thus if the inner product of the vectors $(s^1 - (1/N)1_N)$ and γ is not equal to 0, $\ln P_{GL}^t$ and $\ln P_J^t$ will diverge at a *linear rate* as t increases.[51] In a similar fashion, substitute equations (49) into equations (47) and we obtain the following equations for t = 2,3,...,T:

$$(52)\ \ln P_{GP}^t - \ln P_J^t = [s^t - (1/N)1_N]\cdot[\ln p^t - \ln p^1]$$
$$= [s^1 + \beta(t-1) - (1/N)1_N]\cdot\gamma(t-1)$$
$$= [s^1 - (1/N)1_N]\cdot\gamma(t-1) + \beta\cdot\gamma(t-1)^2.$$

Thus if the inner product of the vectors β and γ is not equal to 0, $\ln P_{GP}^t$ and $\ln P_J^t$ will diverge at a *quadratic rate* as t increases.[52]

Substituting (49) into equations (48) leads to the following equations for t = 2,3,...,T:

[50] We also require that T be small enough so that each $s^t \geq 0_N$.

[51] If the period 1 shares are not all equal and if the rates of growth of log prices differ, then note that we cannot use elementary economic considerations to predict the sign of the inner product $(s^1 - (1/N)1_N)\cdot\gamma$ and thus we cannot predict whether P_J^t will be greater or less than P_{GL}^t.

[52] Again, we cannot predict the sign of the inner product $[s^1 - (1/N)1_N]\cdot\gamma$ but if the products are highly substitutable, it is likely that $\beta\cdot\gamma < 0$. Hence if t is large enough, the quadratic term $\beta\cdot\gamma(t-1)^2$ on the right hand side of (52) will dominate the linear term in t−1 and P_{GP}^t will be less than P_J^t under assumptions (49).

$$(53)\ \ln P_T^t - \ln P_J^t = [(½)s^t + (½)s^1 - (1/N)1_N]\cdot[\ln p^t - \ln p^1]$$
$$= [s^1 - (1/N)1_N]\cdot\gamma(t-1) + (½)\beta\cdot\gamma(t-1)^2$$

Thus if the inner product of the vectors β and γ is not equal to 0, $\ln P_T^t$ and $\ln P_J^t$ will diverge at a *quadratic rate* as t increases.

If all prices are growing at the same geometric rate so that $\gamma = \lambda 1_N$ where λ is a scalar, then using (50), we have $\beta\cdot\gamma = \lambda\beta\cdot 1_N = 0$. We also have $[s^1 - (1/N)1_N]\cdot\gamma = \lambda[s^1 - (1/N)1_N]\cdot 1_N = \lambda[1-1] = 0$. Thus if all prices grow at the same geometric rate, $P_J^t = P_{GL}^t = P_{GP}^t = P_T^t$. Thus in order to for P_J^t to diverge from P_T^t, we require *divergent rates* of price growth.

Assume that shares and log prices are trending linearly approximately so that assumptions (49) and (50) hold approximately. If the N products are highly substitutable, then it is likely that β and γ are *negatively correlated* so that $\beta\cdot\gamma < 0$. In this case, if $\beta_n > 0$ (so that the share of product n is growing over time), then it is likely that $\gamma_n - (1/N)\gamma\cdot 1_N < 0$ so that the price of product n is falling relative to the average growth rate of product prices. In this case, the quadratic term in t on the right hand side of (53), $(½)\beta\cdot\gamma(t-1)^2$, will eventually dominate the linear term in t, $[s^1 - (1/N)1_N]\cdot\gamma(t-1)$, and so for large t, P_T^t will tend to be less than P_J^t. If the products are not close substitutes, then it is likely that $\beta\cdot\gamma > 0$ and hence for large t, P_T^t will tend to be greater than P_J^t. If t is small, then we cannot be sure that the quadratic term in $t-1$ will dominate the linear term on the right hand side of (53) and thus we cannot assert that P_T^t will tend to be greater than P_J^t, since the sign of the linear term is uncertain.

The results in this section can be summarized as follows: the unweighted Jevons (and Dutot) indexes, P_J^t and P_D^t, can provide a reasonable approximation to a fixed base superlative index like P_T^t provided that the expenditure shares do not systematically trend with time and prices do not systematically grow at diverging rates. If these assumptions are not satisfied, then it is likely that the Jevons index will have some substitution bias; P_J^t is likely to exceed P_T^t as t becomes large if the products are close substitutes[53] and P_J^t is likely to be less than P_T^t as t becomes large if the products are not close substitutes.

6. Relationships between Superlative Fixed Base Indexes and Geometric Indexes that use Average Annual Shares as Weights

We consider the properties of weighted Jevons indexes where the weight vector is an annual average of observed monthly shares. Recall that the weighted Jevons (or Cobb Douglas) price index $P_{J\alpha}^t$ was defined by (15) in section 2 as $P_{J\alpha}^t \equiv \prod_{n=1}^N (p_{tn}/p_{1n})^{\alpha_n}$ where the product weighting vector α satisfied the restrictions $\alpha >> 0_N$ and

[53] For our empirical example, P_J^t ended up below P_T^t, contrary to our expectations. However, for small t, the linear term in $(t-1)$ on the right hand side of (53) can dominate the quadratic term in $(t-1)^2$.

$\alpha \cdot 1_N = 1$. The following counterparts to the covariance identities (46)-(48) hold for t = 1,...,T where the Geometric Young index or weighted Jevons index $P_{J\alpha}{}^t$ has replaced $P_J{}^t$:[54]

$$(54)\ \ln P_{GL}{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N [s_{1n} - \alpha_n][\ln p_{tn} - \ln p_{1n}]$$
$$= N\mathrm{Cov}(s^1 - \alpha, \ln p^t - \ln p^1);$$
$$(55)\ \ln P_{GP}{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N [s_{tn} - \alpha_n][\ln p_{tn} - \ln p_{1n}]$$
$$= N\mathrm{Cov}(s^t - \alpha, \ln p^t - \ln p^1);$$
$$(56)\ \ln P_T{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N [(½)s_{tn} + (½)s_{1n} - \alpha_n][\ln p_{tn} - \ln p_{1n}]$$
$$= N\mathrm{Cov}[(½)s^t + (½)s^1 - \alpha, \ln p^t - \ln p^1]$$
$$= (½)[\ln P_{GL}{}^t - \ln P_{J\alpha}{}^t] + (½)[\ln P_{GP}{}^t - \ln P_{J\alpha}{}^t].$$

Define α as the *arithmetic average of the first T^* observed share vectors* s^t:

$$(57)\ \alpha \equiv \Sigma_{t=1}{}^{T*} (1/T^*)s^t.$$

In the context where the data consists of monthly periods, T^* will typically be equal to 12; i.e., the elementary index under consideration is the weighted Jevons index $P_{J\alpha}{}^t$ where the weight vector α is the average of the observed expenditure shares for the first 12 months in the sample.[55]

The decompositions (54)-(56) will hold for the α defined by (57). For a large T^*, α will be close to a constant vector. If the N products are highly substitutable, it is likely that $\mathrm{Cov}(s^1 - \alpha, \ln p^t - \ln p^1) > 0$ and $\mathrm{Cov}(s^t - \alpha, \ln p^t - \ln p^1) < 0$ and hence it is likely that $P_{GL}{}^t > P_{J\alpha}{}^t$ and $P_{GP}{}^t < P_{J\alpha}{}^t$. If the products are not close substitutes, then it is likely that $P_{GL}{}^t < P_{J\alpha}{}^t$ and $P_{GP}{}^t > P_{J\alpha}{}^t$. If there are no divergent trends in prices, then it is possible that the *average share price index* $P_{J\alpha}{}^t$ could provide an adequate approximation to the superlative Törnqvist index $P_T{}^t$.

Note that t takes on the values t = 1,...,T in equations (54)-(56). However, annual share indexes that are implemented by statistical agencies are not constructed in exactly this manner. The practical month to month indexes that are constructed by statistical agencies using annual shares of the type defined by (57) do not choose the reference month for prices to be month 1; rather they chose the reference month for prices to be $T^* + 1$, the month that follows the first year.[56] Thus the reference *year* for share weights precedes the reference *month* for prices. In this case, the logarithm of the month $t \geq T^* + 1$ annual share weighted Jevons index, $\ln P_{J\alpha}{}^t$, is defined as follows where α is the vector of annual average share weights defined by (57):

$$(58)\ \ln P_{J\alpha}{}^t \equiv \Sigma_{n=1}{}^N \alpha_n[\ln p_{tn} - \ln p_{T^*+1,n}]\ ; \qquad t = T^*+1, T^*+2, \ldots, T.$$

[54] The relationship (56) was obtained by Armknecht and Silver (2014; 9); i.e., take logarithms on both sides of their equation (12) and we obtain the first equation in equations (56).
[55] For our empirical example, $T^* = 13$ since we aggregated weekly data for a year into 13 "months" where each month consisted of 4 consecutive weeks.
[56] In actual practice, the reference month for prices can be many months after T^*.

The following counterparts to the identities (54)-(56) hold for $t = T^*+1,T^*+2,...,T$ where α is defined by (57) and $P_{J\alpha}{}^t$ is defined by (58):

$$(59)\ \ln P_{GL}{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N\, [s_{T^*+1,n} - \alpha_n][\ln p_{tn} - \ln p_{T^*+1,n}]$$
$$= N\mathrm{Cov}(s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1});$$
$$(60)\ \ln P_{GP}{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N\, [s_{tn} - \alpha_n][\ln p_{tn} - \ln p_{T^*+1,n}]$$
$$= N\mathrm{Cov}(s^t - \alpha, \ln p^t - \ln p^{T^*+1});$$
$$(61)\ \ln P_T{}^t - \ln P_{J\alpha}{}^t = \Sigma_{n=1}{}^N\, [(½)s_{tn} + (½)s_{T^*+1,n} - \alpha_n][\ln p_{tn} - \ln p_{T^*+1,n}]$$
$$= N\mathrm{Cov}[(½)s^t + (½)s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1}]$$
$$= (½)[\ln P_{GL}{}^t - \ln P_{J\alpha}{}^t] + (½)[\ln P_{GP}{}^t - \ln P_{J\alpha}{}^t].$$

If the N products are highly substitutable, it is likely that $\mathrm{Cov}(s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1}) > 0$ so that $P_{GL}{}^t > P_{J\alpha}{}^t$. It is also likely that $\mathrm{Cov}(s^t - \alpha, \ln p^t - \ln p^{T^*+1}) < 0$ and hence it is likely that $P_{GP}{}^t < P_{J\alpha}{}^t$ in the highly substitutable case. If the products are not close substitutes, then it is likely that $P_{GL}{}^t < P_{J\alpha}{}^t$ and $P_{GP}{}^t > P_{J\alpha}{}^t$. If there are no divergent trends in prices, then it is possible that the *average share price index* $P_{J\alpha}{}^t$ could provide an adequate approximation to the superlative Törnqvist index $P_T{}^t$. However, if there are divergent trends in prices and shares and the products are highly substitutable with each other, then we expect the covariance in (60) to be more negative than the covariance in (59) is positive so that $P_T{}^t$ will tend to be less than the annual shares geometric index $P_{J\alpha}{}^t$.[57]

As usual, there are three simple sets of conditions that will imply that $P_T{}^t = P_{J\alpha}{}^t$: (i) the covariance on the right hand side of (61) equals 0; i.e., $\mathrm{Cov}[(½)s^t + (½)s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1}] = 0$ or equivalently, $\mathrm{Cov}(s^{T^*+1} - \alpha, \ln p^t - \ln p^{T^*+1}) = -\,\mathrm{Cov}(s^t - \alpha, \ln p^t - \ln p^{T^*+1})$; (ii) period t price proportionality (to the prices of the price reference period); i.e., $p^t = \lambda_t p^{T^*+1}$ for some $\lambda_t > 0$; (iii) the arithmetic average of the period T^*+1 and t sales shares are all equal to α; i.e., $(½)s^t + (½)s^{T^*+1} = \alpha$. This last condition will hold if the shares s^t are constant over time and α is defined by (57).

Suppose that there are linear trends in shares and divergent linear trends in log prices; i.e., suppose that assumptions (49) and (50) hold. We look at the implications of these assumptions for the exact relationships (54)-(56). Substituting (49) into equations (56) leads to the following equations for $t = 2,3,...,T$:

$$(62)\ \ln P_T{}^t - \ln P_{J\alpha}{}^t = [(½)s^t + (½)s^1 - \alpha]\cdot[\ln p^t - \ln p^1]$$
$$= [s^1 - \alpha]\cdot\gamma(t-1) + (½)\beta\cdot\gamma(t-1)^2$$
$$= -\,(½)\beta\cdot\gamma(T^*-1)(t-1) + (½)\beta\cdot\gamma(t-1)^2 \qquad \text{using (49) and (57)}$$
$$= (½)\beta\cdot\gamma(t-T^*)(t-1).$$

Thus if the inner product of the vectors β and γ is not equal to 0, $\ln P_T{}^t$ and $\ln P_{J\alpha}{}^t$ will diverge at a *quadratic rate* as t increases. Under these trend assumptions, the average

[57] Our frozen juice data were consistent with the products being highly substitutable. Thus for the 26 periods running from "month" 14 to 39, the arithmetic means for the $P_{GP}{}^t$, $P_T{}^t$, $P_{J\alpha}{}^t$, $P_{GL}{}^t$ were 0.81005, 0.83100, 0.83338 and 0.85270 respectively, which is consistent with our a priori expectations.

share geometric index $P_{J\alpha}{}^t$ will be subject to some substitution bias (as compared to $P_T{}^t$ which controls for substitution bias[58]), which will grow over time.[59]

Note that in real life, new products appear and existing products disappear. Our analysis can take this fact into account *in theory*: we are assuming that we have somehow calculated reservation prices for products that are not available in the current period. However, product churn means that shares are not constant over time; i.e., product churn will lead to (nonsmooth) trends in product shares. Superlative indexes like $P_F{}^t$ and $P_T{}^t$ can deal with new and disappearing products in a way that is consistent with consumer theory, provided that suitable reservation prices have been either estimated or approximated by suitable rules of thumb.

7. To Chain or Not to Chain

In the above discussions, we have focussed on direct indexes that compare the prices of period t with the prices of period 1. But it is also possible to move from period 1 prices to period t prices by jumping from one period to the next and cumulating the jumps. If the second method is used, the resulting period t price index is called a *chained index*. In this section, we will examine the possible differences between direct and chained Törnqvist price indexes.

It is convenient to introduce some new notation. Denote the Törnqvist price index that compares the prices of period j to the prices of period i (the base period for the comparison) by $P_T(i,j)$. The logarithm of $P_T(i,j)$ is defined as follows for i,j = 1,...,N:

$$\begin{aligned}(63)\ \ln P_T(i,j) &\equiv (½)\Sigma_{n=1}{}^N (s_{in} + s_{jn})(\ln p_{jn} - \ln p_{in}) \\ &= (½)(s^i + s^j)\cdot(\ln p^j - \ln p^i).\end{aligned}$$

The chained Törnqvist price index going from period 1 to T will coincide with the corresponding direct index if the indexes $P_T(i,j)$ satisfy the following *multiperiod identity test* which is due to Walsh (1901; 389), (1921; 540):

$$(64)\ P_T(1,2)P_T(2,3)\ldots P_T(T-1,T)P_T(T,1) = 1.$$

The above test can be used to measure the amount that the chained indexes between periods 1 and T differ from the corresponding direct index that compares the prices of period 1 and T; i.e., if the product of indexes on the left hand side of if (64) is different from unity, then we say that the index number formula is subject to *chain drift* and the difference between the left and right hand sides of (64) serves to measure the magnitude of the chain drift problem.[60] In order to determine whether the Törnqvist price index

[58] See Diewert (1976) on this point.

[59] As was the case in the previous section, if all prices grow at the same geometric rate, then $P_{J\alpha}{}^t = P_{GL}{}^t = P_{GP}{}^t = P_T{}^t$.

[60] Walsh (1901; 401) was the first to propose this methodology to measure chain drift. It was independently proposed later by Persons (1921; 110) and Szulc (1983; 540). Fisher's (1922; 284) circular gap test could also be interpreted as a test for chain drift.

formula satisfies the multiperiod identity test (64), take the logarithm of the left hand side of (64) and check whether it is equal to the logarithm of 1 which is 0. Thus substituting definitions (63) into the logarithm of the left hand side of (64) leads to the following expressions:[61]

$$
\begin{aligned}
(65)\ & \ln P_T(1,2) + \ln P_T(2,3) + \ldots + \ln P_T(T-1,T) + \ln P_T(T,1) \\
&= \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{1n}+s_{2n})(\ln p_{2n}-\ln p_{1n}) + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{2n}+s_{3n})(\ln p_{3n}-\ln p_{2n}) + \ldots \\
&\quad + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{T-1,n}+s_{Tn})(\ln p_{Tn}-\ln p_{T-1,n}) + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{Tn}+s_{1n})(\ln p_{1n}-\ln p_{Tn}) \\
&= \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{1n}-s_{3n})\ln p_{2n} + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{2n}-s_{4n})\ln p_{3n} + \ldots + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{T-2,n}-s_{Tn})\ln p_{T-1,n} \\
&\quad + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{Tn}-s_{2n})\ln p_{1n} + \tfrac{1}{2}\Sigma_{n=1}^{N} (s_{T-1,n}-s_{1n})\ln p_{Tn}.
\end{aligned}
$$

In general, it can be seen that the Törnqvist price index formula will be subject to some chain drift i.e., the sums of terms on the right hand side of (65) will not equal 0 in general. However there are four sets of conditions where these terms will sum to 0.

The first set of conditions makes use of the first equality on the right hand side of (65). If the prices vary in strict proportion over time, so that $p^t = \lambda_t p^1$ for $t = 2,3,...,T$, then it is straightforward to show that (64) is satisfied.

The second set of conditions makes use of the second equality in equations (65). If the shares s^t are constant over time,[62] then it is obvious that (64) is satisfied.

The third set of conditions also makes use of the second equality in (65). The sum of terms $\Sigma_{n=1}^{N} (s_{1n}-s_{3n})\ln p_{2n}$ is equal to $(s^1-s^3)\cdot \ln p^2$ which in turn is equal to $(s^1-s^3)\cdot(\ln p^2-\ln p^{2*}) = N\mathrm{Cov}(s^1-s^3,\ln p^2)$ where $\ln p^{2*} \equiv (1/N)\ \Sigma_{n=1}^{N} \ln p_{2n}$, the mean of the components of $\ln p^2$. Thus the N sets of summations on the right hand side of the second equation in (65) can be interpreted as constants time the covariances of a difference in shares (separated by one or more time periods) with the logarithm of a price vector for a time period that is not equal to either of the shares in the difference in shares. Thus if $\mathrm{Cov}(s^1-s^3,\ln p^2) = \mathrm{Cov}(s^2-s^4,\ln p^3) = \ldots = \mathrm{Cov}(s^{T-2}-s^T,\ln p^{T-1}) = \mathrm{Cov}(s^T-s^2,\ln p^1) = \mathrm{Cov}(s^{T-1}-s^1,\ln p^T) = 0$, then (64) will be satisfied. These zero covariance conditions will be satisfied if the log prices of one period are uncorrelated with the shares of all other periods. If the time period is long enough and there are no trends in log prices and shares, so that prices are merely bouncing around in a random fashion,[63] then these zero covariance conditions are likely to be satisfied to a high degree of approximation and thus under these conditions, the Törnqvist Theil price index is likely to be largely free of chain drift. However, in the elementary index context where retailers have periodic highly discounted prices, the zero correlation conditions are unlikely to hold. Suppose that product n goes on sale during period 2 so that $\ln p_{2n}$ is well below the average price for period 2. Suppose product n is not on sale during periods 1 and 3. If purchasers have stocked up on product n during period t, it is likely that s_{3n} will be less than s_{1n} and thus it

[61] Persons (1928; 101) developed a similar decomposition using the bilateral Fisher formula instead of the Törnqvist formula.
[62] If purchasers of the products have Cobb-Douglas preferences, then the sales shares will be constant.
[63] Szulc (1983) introduced the term "price bouncing" to describe the behavior of soft drink prices in Canada at the elementary level.

is likely that $Cov(s^1–s^3,lnp^2) < 0$. Now suppose that product n is not on sale during period 2. In this case, it is likely that lnp_{2n} is greater than the average log price during period 2. If product n was on sale during period 1 but not period 3, then s_{1n} will tend to be greater than s_{3n} and thus $Cov(s^1–s^3,lnp^2) > 0$. However, if product n was on sale during period 3 but not period 1, then s_{1n} will tend to be less than s_{3n} and thus $Cov(s^1–s^3,lnp^2) < 0$. These last two cases should largely offset each other and so we are left with the likelihood that $Cov(s^1–s^3,lnp^2) < 0$. Similar arguments apply to the other covariances and so we are left with the expectation that the chained Törnqvist index used in the elementary index context is likely to drift downwards relative to its fixed base counterpart.[64]

Since the Fisher index normally approximates the Törnqvist fairly closely, we expect both the chained Fisher and Törnqvist indexes to exhibit downward chain drift. However, in our empirical example, the above expectation that the Fisher index is subject to downward chain drift was not realized: there was substantial upward chain drift for both indexes. Feenstra and Shapiro (2003) also found upward chain drift in the Törnqvist formula using an alternative scanner data set. Persons (1928; 100-105) had an extensive discussion of the chain drift problem with the Fisher index and he gave a numerical example on page 102 of his article which showed how upward chain drift could occur. We have adapted his example in Table 1 below.

Table 1: Prices and Quantities for Two Products and the Fisher Fixed Base and Chained Price Indexes

t	p_1^t	p_2^t	q_1^t	q_2^t	P_F^t	P_{FCh}^t
1	2	1	100	1	1.00000	1.00000
2	10	1	40	40	4.27321	4.27321
3	10	1	25	80	3.55553	4.27321
4	5	2	50	20	2.45676	2.96563

Product 1 is on sale in period 1 and goes back to a relatively high price in periods 2 and 3 and then goes on sale again but the discount is not as steep as the period 1 discount. Product 2 is at its "regular" price for periods 1-3 and then rises steeply in period 4. Products 1 and 2 are close substitutes so when product 1 is steeply discounted, only 1 unit of product 2 is sold in period 1 while 100 units of product 1 are sold. When the price of product 1 increases fivefold in period 2, demand for the product falls and purchasers switch to product 2 but the adjustment to the new higher price of product 1 is not complete in period 2: in period 3 (where prices are unchanged from period 2), purchasers continue to substitute away from product 1 and towards product 2. It is this incomplete adjustment that causes the chained index to climb above the fixed base index in period

[64] Fisher (1922; 284) found little difference in the fixed base and chained Fisher indexes for his particular data set which he used to compare 119 different index number formulae. Fisher (1922; noted that the Carli, Laspeyres and share weighted Carli chained indexes showed upward chain drift. However, Persons (1921; 110) showed that the Fisher chained index ended up about 4% lower than its fixed base counterpart for his agricultural data set covering 10 years. This is an early example of the downward chain drift associated with the use of the Fisher index.

3.[65] Thus it is not always the case that the Fisher index is subject to downward chain drift but we do expect that "normally", this would be the case.

The fourth set of conditions that ensure that there is no chain drift are assumptions (49) and (50); i.e., the assumption that *shares and log prices have linear trends*. To prove this assertion, substitute equations (49) into either one of the two right hand side equations in (65) and we find that the resulting sum of terms is 0.[66] This result is of some importance at higher levels of aggregation where aggregate prices and quantities are more likely to have smooth trends. If the trends are actually linear, then this result shows that there will be no chain drift if the Törnqvist Theil index number formula is used to aggregate the data.[67] However, when this formula is used at the elementary level when there are frequent fluctuations in prices and quantities, chain drift is likely to occur and thus the use of a fixed base index or a multilateral index is preferred under these conditions.

The main advantage of the chain system is that under conditions where prices and quantities are trending smoothly, chaining will reduce the spread between the Paasche and Laspeyres indexes.[68] These two indexes each provide an asymmetric perspective on the amount of price change that has occurred between the two periods under consideration and it could be expected that a single point estimate of the aggregate price change should lie between these two estimates. Thus at higher levels of aggregation, the use of either a chained Paasche or Laspeyres index will usually lead to a smaller difference between the two and hence to estimates that are closer to the "truth".

Hill (1988; 136-137), drawing on the earlier research of Szulc (1983), noted that it is not appropriate to use the chain system when prices oscillate (or "bounce" to use Szulc's (1983; 548) term). This phenomenon can occur in the context of regular seasonal fluctuations or in the context of periodic heavily discounting of prices. However, in the context of roughly monotonically changing prices and quantities, Hill recommended the use of chained symmetrically weighted indexes.[69] The Fisher and Törnqvist price index indexes are examples of symmetrically weighted indexes.

An alternative to the use of a fixed base index is the use of a *multilateral index*. A problem with the use of a fixed base index is that it depends asymmetrically on the choice of the base period. If the structure of prices and quantities for the base period is unusual and fixed base index numbers are used, then the choice of the base period could lead to "unusual" results. Multilateral indexes treat each period symmetrically and thus avoid

[65] Persons (1928; 102) explained that it was *incomplete adjustment* that caused the Fisher chained index to climb above the corresponding fixed base index in his example.
[66] This result was first established by Alterman, Diewert and Feenstra (1999; 61-65).
[67] This transitivity property carries over to an *approximate* transitivity property for the Fisher and Walsh index number formulae using the fact that these indexes approximate the Törnqvist Theil index to the second order around an equal price and quantity point; see Diewert (1978) on these approximations.
[68] See Diewert (1978; 895) and Hill (1988) for additional discussion on the benefits and costs of chaining.
[69] Vartia and Suoperä (2018) also recommend the use of symmetrically weighted index numbers.

this problem. In the following section, we will introduce some possible multilateral indexes that are free of chain drift (within our window of T observations).[70]

8. Relationships between the Törnqvist Index and the GEKS and CCDI Multilateral Indexes

It is useful to introduce some additional notation at this point. Denote the Laspeyres, Paasche and Fisher price indexes that compare the prices of period j to the prices of period i (the base period for the comparison) by $P_L(i,j)$, $P_P(i,j)$ and $P_F(i,j)$ respectively. These indexes are defined as follows for r,t = 1,...,N:

(66) $P_L(r,t) \equiv p^t \cdot q^r / p^r \cdot q^r$;
(67) $P_P(r,t) \equiv p^t \cdot q^t / p^r \cdot q^t$;
(68) $P_F(r,t) \equiv [P_L(r,t)P_P(r,t)]^{1/2}$.

The Fisher indexes have very good axiomatic properties and hence are preferred indexes from the viewpoint of the test or axiomatic approach.[71]

Obviously, one could choose period 1 as the base period and form the following sequence of price levels relative to period 1: $P_F(1,1) = 1$, $P_F(1,2)$, $P_F(1,3)$, ..., $P_F(1,T)$. But one could also use period 2 as the base period and use the following sequence of price levels: $P_F(2,1)$, $P_F(2,2) = 1$, $P_F(2,3)$, ..., $P_F(2,T)$. Each period could be chosen as the base period and thus we end up with T alternative series of Fisher price levels. Since each of these sequences of price levels is equally plausible, Gini (1931) suggested that it would be appropriate to take the geometric average of these alternative price levels in order to determine the final set of price levels. Thus the *GEKS price levels*[72] for periods t = 1,2,...,T are defined as follows:

(69) $p_{GEKS}^t \equiv [\prod_{r=1}^T P_F(r,t)]^{1/T}$.

Note that all time periods are treated in a symmetric manner in the above definitions. The GEKS price indexes P_{GEKS}^t are obtained by normalizing the above price levels so that the period 1 index is equal to 1. Thus we have the following definitions for P_{GEKS}^t for t = 1,...,T:

(70) $P_{GEKS}^t \equiv p_{GEKS}^t / p_{GEKS}^1$.

It is straightforward to verify that the GEKS price indexes satisfy Walsh's multiperiod identity test which becomes the following test in the present context:

---

[70] Ivancic, Diewert and Fox (2009) (2011) advocated the use of multilateral indexes adapted to the time series context in order to control chain drift. Balk (1980) (1981) also advocated the use of multilateral indexes in order to address the problem of seasonal commodities.
[71] See Diewert (1992) on the axiomatic properties of the Fisher index.
[72] Eltetö and Köves (1964) and Szulc (1964) independently derived the GEKS price indexes by an alternative route. Thus the name GEKS has the initials of all four primary authors of the method. Ivancic, Diewert and Fox (2009) (2011) suggested the use of the GEKS index in the time series context.

(71) $[P_{GEKS}^2/P_{GEKS}^1][P_{GEKS}^3/P_{GEKS}^2]...[P_{GEKS}^T/P_{GEKS}^{T-1}][P_{GEKS}^1/P_{GEKS}^T] = 1$.

Thus the GEKS indexes are not subject to chain drift within the window of T periods under consideration.

Recall definition (63) which defined the logarithm of the Törnqvist price index, $\ln P_T(i,j)$, that compared the prices of period j to the prices of period i. The GEKS methodology can be applied using $P_T(r,t)$ in place of the Fisher $P_F(r,t)$ as the basic bilateral index building block. Thus define the *period t GEKS Törnqvist price level*, p_{GEKST}^t, for t = 1,...,T as follows:

(72) $p_{GEKST}^t \equiv [\prod_{r=1}^T P_T(r,t)]^{1/T}$.

The *GEKST price indexes* P_{GEKST}^t are obtained by normalizing the above price levels so that the period 1 index is equal to 1. Thus we have the following definitions for P_{GEKST}^t for t = 1,...,T:

(73) $P_{GEKST}^t \equiv p_{GEKST}^t/p_{GEKST}^1$.

Since $P_T(r,t)$ approximates $P_F(r,t)$ to the second order around an equal price and quantity point, the P_{GEKST}^t will usually be quite close to the corresponding P_{GEKS}^t indexes.

It is possible to provide a very simple alternative approach to the derivation of the GEKS Törnqvist price indexes.[73] Define the *sample average sales share* for product n, $s_{\bullet n}$, and the *sample average log price* for product n, $\ln p_{\bullet n}$, as follows for n = 1,...,N:

(74) $s_{\bullet n} \equiv \Sigma_{t=1}^T (1/T)s_{tn}$;
(75) $\ln p_{\bullet n} \equiv \Sigma_{t=1}^T (1/T)\ln p_{tn}$.

The logarithm of the *CCDI price level for period t*, $\ln p_{CCDI}^t$, is defined by comparing the prices of period t with the sample average prices using the bilateral Törnqvist formula; i.e., for t = 1,...,T, we have the following definitions:

(76) $\ln p_{CCDI}^t \equiv \Sigma_{n=1}^N \frac{1}{2}(s_{tn} + s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n})$.

The *CCDI price index for period t*, P_{CCDI}^t, is defined as the following normalized CCDI price level for t = 1,...,T:

(77) $P_{CCDI}^t \equiv p_{CCDI}^t/p_{CCDI}^1$.

Using the above definitions, the logarithm of the CCDI price index for period t is equal to the following expressions for t = 1,...,T:

[73] This approach is due to Inklaar and Diewert (2015). It is an adaptation of the distance function approach used by Caves, Christensen and Diewert (1982) to the price index context.

$$(78)\ \ln P_{CCDI}{}^t = \ln p_{CCDI}{}^t - \ln p_{CCDI}{}^1$$
$$= \Sigma_{n=1}^N (½)(s_{tn} + s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n}) - \Sigma_{n=1}^N (½)(s_{1n} + s_{\bullet n})(\ln p_{1n} - \ln p_{\bullet n})$$
$$= \ln P_T{}^t + \Sigma_{n=1}^N (½)(s_{tn} - s_{\bullet n})(\ln p_{1n} - \ln p_{\bullet n}) - \Sigma_{n=1}^N (½)(s_{1n} - s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n})$$
$$= \ln P_{GEKST}{}^t$$

where the last equality follows by direct computation or by using the computations in Inklaar and Diewert (2015).[74] Thus the CCDI multilateral price indexes are equal to the GEKS Törnqvist multilateral indexes defined by (73). Define $s^\bullet \equiv [s_{\bullet 1},..., s_{\bullet N}]$ as the vector of sample average shares and $\ln p^\bullet \equiv [\ln p_{\bullet 1},...,\ln p_{\bullet N}]$ as the vector of sample average log prices. Then the last two terms on the right hand side of the penultimate equality in (80) can be written as $(½)NCov(s^t - s^\bullet,\ln p^1 - \ln p^\bullet) - (½)NCov(s^1 - s^\bullet,\ln p^t - \ln p^\bullet)$. For our empirical example, the sample average of these two sets of covariance terms turned out to be 0 with variances equal to 0.00024 and 0.00036 respectively. Thus it is likely that $\ln P_{CCDI}{}^t \approx \ln P_T{}^t$ for each t. Moreover, under the assumptions of linear trends in log prices and linear trends in shares, assumptions (49) and (50), Alterman, Diewert and Feenstra (1999) showed that the period t bilateral Törnqvist price index, $P_T{}^t$, is equal to its chained counterpart for any t.[75] This result implies that $P_T{}^t = P_{CCDI}{}^t = P_{GEKST}{}^t$ for t = 1,...,T under the linear trends assumption. Thus we expect the period t multilateral index, $P_{GEKST}{}^t = P_{CCDI}{}^t$ to approximate the corresponding fixed base period t Törnqvist price index, $P_T{}^t$, provided that prices and quantities have smooth trends.

Since $P_F{}^t$ approximates $P_T{}^t$, we expect that the following approximate equalities will hold under the smooth trends assumption for t = 1,...,T:

$$(79)\ P_F{}^t \approx P_T{}^t \approx P_{GEKS}{}^t \approx P_{GEKST}{}^t = P_{CCDI}{}^t.$$

The above indexes will be free from chain drift within the window of T periods;[76] i.e., if prices and quantities for any two periods in the sample are equal, then the price index will register the same value for these two periods.

9. Unit Value Price and Quantity Indexes

As was mentioned in section 2, there is a preliminary aggregation over time problem that needs to be addressed; i.e., exactly how should the period t prices and quantities for commodity n, $p_n{}^t$ and $q_n{}^t$, that are used in an index number formula be defined? During any time period t, there will typically be many transactions in a specific commodity n at a number of different prices. Hence, there is a need to provide a more precise definition for the "average" or "representative" price for commodity n in period t, $p_n{}^t$. Starting with

[74] The second from last equality was derived in Diewert and Fox (2017; 17).
[75] See the discussion below equation (65). Note that the assumption of linear trends in shares is not consistent with the existence of new and disappearing products.
[76] See de Haan (2015) and Diewert and Fox (2017) for discussions of the problems associated with linking the results from one rolling window multilateral comparison to a subsequent window of observations. Empirically, there does not appear to be much chain drift between the indexes generated by subsequent windows.

Drobisch (1871), many measurement economists and statisticians advocated the use of the *unit value* (total value transacted divided by total quantity) as the appropriate price p_n^t for commodity n and the total quantity transacted during period t as the appropriate quantity, q_n^t; e.g., see Walsh (1901; 96) (1921; 88), Fisher (1922; 318) and Davies (1924; 183) (1932; 59). If it is desirable to have q_n^t be equal the total quantity of commodity n transacted during period t and also desirable to have the product of the price p_n^t times quantity q_n^t to be equal the value of period t transactions in commodity n, then one is *forced* to define the aggregate period t price for commodity n, p_n^t, to be the total value transacted during the period divided by the total quantity transacted, which is the unit value for commodity n.[77]

There is general agreement that a unit value price is an appropriate price concept to be used in an index number formula if the transactions refer to a narrowly defined homogeneous commodity. Our task in this section is to look at the properties of a unit value price index when aggregating over commodities that are not completely homogeneous. We will also look at the properties of the companion quantity index in this section.

The period t *unit value price level*, p_{UV}^t, and the corresponding period t *unit value price index*, P_{UV}^t, are defined as follows for t = 1,...,T:

$$(80)\ p_{UV}^t \equiv p^t \cdot q^t / 1_N \cdot q^t \ ;$$
$$\begin{aligned}(81)\ P_{UV}^t &\equiv p_{UV}^t / p_{UV}^1 \\ &= [p^t \cdot q^t / 1_N \cdot q^t]/[p^1 \cdot q^1 / 1_N \cdot q^1] \\ &= [p^t \cdot q^t / p^1 \cdot q^1]/Q_{UV}^t\end{aligned}$$

where the period t *unit value quantity index*, Q_{UV}^t, is defined as follows for t = 1,...,T:

$$(82)\ Q_{UV}^t \equiv 1_N \cdot q^t / 1_N \cdot q^1 .$$

It can be seen that the unit value price index satisfies Walsh's multiperiod identity test and thus P_{UV}^t is free of chain drift.[78]

We will look at the relationship of the *unit value quantity indexes*, Q_{UV}^t, with the corresponding *Laspeyres*, *Paasche* and *Fisher fixed base quantity indexes*, Q_L^t, Q_P^t and Q_F^t, defined below for t = 1,...,T:

$$(83)\ Q_L^t \equiv p^1 \cdot q^t / p^1 \cdot q^1 = \Sigma_{n=1}^N s_{1n}(q_{tn}/q_{1n}) \ ;$$
$$(84)\ Q_P^t \equiv p^t \cdot q^t / p^t \cdot q^1 = [\Sigma_{n=1}^N s_{tn}(q_{tn}/q_{1n})^{-1}]^{-1} \ ;$$
$$(85)\ Q_F^t \equiv [Q_L^t Q_P^t]^{1/2} .$$

[77] For additional discussion on unit value price indexes, see Balk (2008; 72-74) and Diewert and von der Lippe (2010).

For the second set of equations in (83), we require that $q_{1n} > 0$ for all n and for the second set of equations in (84), we require that all $q_{tn} > 0$. Recall that the period t *sales* share vector $s^t \equiv [s_{t1},...,s_{tN}]$ was defined at the beginning of section 2. The period t *quantity* share vector $S^t \equiv [S_{t1},...,S_{tN}]$ was also defined in section 2 as follows for t = 1,...,T:

$$(86)\ S^t \equiv q^t/1_N{\cdot}q^t .$$

Below, we will make use of the following identities (87), which hold for t = 1,...,T:

$$\begin{aligned}(87)\ \Sigma_{n=1}^N [p_{UV}^t - p_{tn}]q_{tn} &= \Sigma_{n=1}^N [(p^t{\cdot}q^t/1_N{\cdot}q^t) - p_{tn}]q_{tn} && \text{using definitions (80)}\\ &= (p^t{\cdot}q^t/1_N{\cdot}q^t)1_N{\cdot}q^t - p^t{\cdot}q^t\\ &= 0.\end{aligned}$$

The following relationships between Q_{UV}^t and Q_L^t hold for t = 1,...,T:

$$\begin{aligned}(88)\ Q_{UV}^t - Q_L^t &= [1_N{\cdot}q^t/1_N{\cdot}q^1] - [p^1{\cdot}q^t/p^1{\cdot}q^1] && \text{using (82) and (83)}\\ &= \Sigma_{n=1}^N S_{1n}(q_{tn}/q_{1n}) - \Sigma_{n=1}^N s_{1n}(q_{tn}/q_{1n}) && \text{using (86) and (83)}\\ &= \Sigma_{n=1}^N [S_{1n} - s_{1n}](q_{tn}/q_{1n})\\ &= N\mathrm{Cov}(S^1 - s^1,q^t/q^1)\end{aligned}$$

where the vector of period t to period 1 relative quantities is defined as $q^t/q^1 \equiv [q_{t1}/q_{11}, q_{t2}/q_{12}...,q_{tn}/q_{1N}]$. As usual, there are 3 special cases of (88) which will imply that $Q_{UV}^t = Q_L^t$. (i) $S^1 = s^1$ so that the vector of period 1 real quantity shares S^1 is equal to the period 1 sales share vector s^1. This condition is equivalent to $p^1 = \lambda_1 1_N$ so that all period 1 prices are equal. (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time. (iii) $\mathrm{Cov}(S^1 - s^1,q^t/q^1) = 0$.[79]

There are two problems with the above bias formula: (i) it is difficult to form a judgement on the sign of the covariance $\mathrm{Cov}(S^1 - s^1,q^t/q^1)$ and (ii) the decomposition given by (88) requires that all components of the period 1 quantity vector be positive.[80] It would be useful to have a decomposition that allowed some quantities (and sales shares) to be equal to 0. Consider the following alternative decomposition to (89) for t = 1,...,T:

$$\begin{aligned}(89)\ Q_{UV}^t - Q_L^t &= [1_N{\cdot}q^t/1_N{\cdot}q^1] - [p^1{\cdot}q^t/p^1{\cdot}q^1] && \text{using (82) and (83)}\\ &= \Sigma_{n=1}^N [(q_{tn}/1_N{\cdot}q^1) - (p_{1n}q_{tn}/p^1{\cdot}q^1)]\\ &= \Sigma_{n=1}^N [(1/1_N{\cdot}q^1) - (p_{1n}/p^1{\cdot}q^1)]q_{tn}\\ &= \Sigma_{n=1}^N [(p^1{\cdot}q^1/1_N{\cdot}q^1) - p_{1n}][q_{tn}/p^1{\cdot}q^1]\\ &= \Sigma_{n=1}^N [p_{UV}^1 - p_{1n}][q_{tn}/p^1{\cdot}q^1] && \text{using (80) for t = 1}\\ &= \Sigma_{n=1}^N [p_{UV}^1 - p_{1n}][q_{tn} - q_{1n}Q_{UV}^t]/p^1{\cdot}q^1 && \text{using (87) for t = 1}\\ &= Q_{UV}^t\, \Sigma_{n=1}^N [p_{UV}^1 - p_{1n}][(q_{tn}/Q_{UV}^t) - q_{1n}]/p^1{\cdot}q^1\\ &= Q_{UV}^t\, \Sigma_{n=1}^N s_{1n}[(p_{UV}^1/p_{1n}) - 1][(q_{tn}/q_{1n}Q_{UV}^t) - 1] && \text{if } q_{1n} > 0 \text{ for all n}\end{aligned}$$

[79] For similar bias formulae, see Balk (2008; 73-74) and Diewert and von der Lippe (2010).

[80] We are assuming that all prices are positive in all periods (so if there are missing prices they must be replaced by positive imputed prices) but we are not assuming that all quantities (and expenditure shares) are positive.

$$= Q_{UV}^t \varepsilon_L^t$$

where the *period t error term* ε_L^t is defined for t = 1,...,T as:

(90) $\varepsilon_L^t \equiv \Sigma_{n=1}^N [p_{UV}^1 - p_{1n}][(q_{tn}/Q_{UV}^t) - q_{1n}]/p^1 \cdot q^1$.[81]

If $q_{1n} > 0$ for n = 1,...,N, then ε_L^t is equal to $\Sigma_{n=1}^N s_{1n}[(p_{UV}^1/p_{1n}) - 1][(q_{tn}/q_{1n}Q_{UV}^t) - 1]$.

Note that the terms on the right hand side of (89) can be interpreted as $(N/p^1 \cdot q^1)$ times the covariance $\mathrm{Cov}(p_{UV}^1 1_N - p^1, q^t - Q_{UV}^t q^1)$ since $1_N \cdot (q^t - Q_{UV}^t q^1) = 0$. If the products are substitutes, it is likely that this covariance is *negative*, since if p_{1n} is unusually low, we would expect that it would be less than the period 1 unit value price level p_{UV}^1 so that $p_{UV}^1 - p_{1n} > 0$. Furthermore, if p_{1n} is unusually low, then we would expect that the corresponding q_{1n} is unusually high, and thus it is likely that q_{1n} is greater than q_{tn}/Q_{UV}^t and so $q_{tn} - q_{1n}Q_{UV}^t < 0$. Thus the N terms in the covariance will tend to be negative provided that there is some degree of substitutability between the products.[82] However, looking at formula (90) for ε_L^t, it can be seen that all terms on the right hand side of (90) do not depend on t, except for the N period t deflated product quantity terms, q_{tn}/Q_{UV}^t for n = 1,...,N. Hence if there is a great deal of variation in the period t quantities q_{tn}, then $q_{tn}/Q_{UV}^t - q_{1n}$ could be positive or negative and thus the tendency for ε_L^t to be negative will be a weak one. Thus our expectation is that the error term ε_L^t is likely to be negative and hence $Q_{UV}^t < Q_L^t$ for $t \geq 2$ but this expectation is a weak one.

Our scanner data set listed in the Appendix uses estimated reservation prices for products that were absent in any period. Note that the period t unit value price and quantity indexes, P_{UV}^t and Q_{UV}^t defined by (81) and (82) above are well defined using reservation prices and 0 quantities for the unavailable products. However, P_{UV}^t and Q_{UV}^t *do not depend on the estimated reservation prices*; i.e., the definitions of P_{UV}^t and Q_{UV}^t zero out the estimated reservation prices. However, the error term ε_L^t defined above by (90) does not zero out the estimated reservation prices for products that are absent in period 1 but present in period t. This makes sense, since $Q_L^t \equiv p^1 \cdot q^t / p^1 \cdot q^t$ will depend on the products n that are absent in period 1 but are present in period t and in defining ε_L^t, we are comparing Q_{UV}^t (does not depend on p_{1n}) to Q_L^t (does depend on p_{1n}). Thus a unit value price index in general *cannot be consistent* with the (Hicksian) economic approach to index number theory if there are new or disappearing products in the sample of products. This same point applies to the use of several multilateral indexes in the context of changes in the availability of products as we shall see later.[83]

---

[81] Note that this error term is homogeneous of degree 0 in the components of p^1, q^1 and q^t. Hence it is invariant to proportional changes in the components of these vectors.

[82] The results in previous sections looked at responses of product *shares* to changes in prices and with data that are consistent with CES preferences, the results depended on whether the elasticity of substitution was greater or less than unity. In the present section, the results depend on whether the elasticity of substitution is equal to 0 or greater than 0; i.e., it is the response of *quantities* (rather than *shares*) to lower prices that matters.

[83] An examination of formula (90) shows that if there are many missing products in period 1, this will tend to increase the probability that ε_L^t is negative.

As usual, there are 3 special cases of (89) which will imply that $Q_{UV}^t = Q_L^t$: (i) $p^1 = \lambda_1 1_N$ so that all period 1 prices are equal; (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time; (iii) $\mathrm{Cov}(p_{UV}^1 1_N - p^1, q^t - Q_{UV}^t q^1) = 0$. These conditions are equivalent to our earlier conditions listed below (88).

If we divide both sides of equation t in equations (89) by Q_{UV}^t, we obtain the following system of identities for t = 1,...,T:

(91) $Q_L^t/Q_{UV}^t = 1 - \varepsilon_L^t$

where we expect ε_L^t to be a small negative number in the elementary index context.

The identities in (89) and (91) are valid if we interchange prices and quantities. The quantity counterparts to p_{UV}^t and P_{UV}^t defined by (80) and (81) are the period t *Dutot quantity level* q_D^t and *quantity index* Q_D^t [84] defined as $q_D^t \equiv p^t \cdot q^t/1_N \cdot p^t = \alpha^t \cdot q^t$ (where $\alpha^t \equiv p^t/1_N \cdot p^t$ is a vector of period t price weights for q^t) and $Q_D^t \equiv q_{UV}^t/q_{UV}^1 = [p^t \cdot q^t/p^1 \cdot q^1]/P_D^t$ where we redefine the period t Dutot price level as $p_D^t \equiv 1_N \cdot p^t$ and the period t Dutot price index as $P_D^t \equiv p_D^t/p_D^1 = 1_N \cdot p^t/1_N \cdot p^1$ which coincides with our earlier definition (10) for P_D^t. Using these definitions and interchanging prices and quantities, equations (91) become the following equations for t = 1,...,T:

(92) $P_L^t/P_D^t = 1 - \varepsilon_L^{t*}$

where the period t error term ε_L^{t*} is defined for t = 1,...,T as:

(93) $\varepsilon_L^{t*} \equiv \Sigma_{n=1}^N [q_D^1 - q_{1n}][(p_{tn}/P_D^t) - p_{1n}]/p^1 \cdot q^1$.

If p_{1n} is unusually low, then it is likely that it will be less than p_{tn}/P_D^t and it is also likely that q_{1n} will be unusually high and hence greater than the average period 1 Dutot quantity level, q_D^1. Thus the N terms in the definition of ε_L^{t*} will tend to be negative and thus $1 - \varepsilon_L^{t*}$ will tend to be greater than 1. Thus there will be a tendency for $P_D^t < P_L^t$ for $t \geq 2$ but again, this expectation is a weak one.

It can be verified that the following identities hold for the period t Laspeyres, Paasche and unit value price and quantity indexes for t = 1,...,T:

(94) $p^t \cdot q^t/p^1 \cdot q^1 = P_{UV}^t Q_{UV}^t = P_P^t Q_L^t = P_L^t Q_P^t$.

Equations (94) imply the following identities for t = 1,...,T:

(95) $P_{UV}^t/P_P^t = Q_L^t/Q_{UV}^t$
$\qquad = 1 - \varepsilon_L^t$

[84] Balk (2008; 7) called Q_{UV}^t a Dutot-type quantity index.

where the last set of equations follow from equations (91). Thus we expect that $P_{UV}{}^t > P_P{}^t$ for t = 2,3,...,T if the products are substitutes and $\varepsilon_L{}^t$ is negative.

We now turn our attention to developing an exact relationship between $Q_{UV}{}^t$ and the Paasche quantity index $Q_P{}^t$. Using definitions (82) and (84), we have for t = 1,...,T:

$$
\begin{aligned}
(96)\ [Q_{UV}{}^t]^{-1} - [Q_P{}^t]^{-1} &= [1_N \cdot q^1/1_N \cdot q^t] - [p^t \cdot q^1/p^t \cdot q^t] && \text{using (82) and (84)} \\
&= \Sigma_{n=1}{}^N [S_{tn} - s_{tn}][q_{1n}/q_{tn}] \\
&= N\mathrm{Cov}(S^t - s^t, q^1/q^t)
\end{aligned}
$$

where the second set of equalities in (96) follows using (88) and (86), assuming that $q_{tn} > 0$ for n = 1,...,N.

As usual, there are 3 special cases of (96) which will imply that $Q_{UV}{}^t = Q_P{}^t$: (i) $S^t = s^t$ so that the vector of period t real quantity shares S^t is equal to the period t sales share vector s^t. This condition is equivalent to $p^t = \lambda_t 1_N$ so that all period t prices are equal. (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time. (iii) $N\mathrm{Cov}(S^t - s^t, q^1/q^t) = 0$.

Again, there are two problems with the above bias formula: (i) it is difficult to form a judgement on the sign of the covariance $N\mathrm{Cov}(S^t - s^t, q^1/q^t)$ and (ii) the decomposition given by (96) requires that all components of the period t quantity vector be positive. We will proceed to develop a decomposition that does not require the positivity of q^t. The following exact decomposition holds for t = 1,...,T:

$$
\begin{aligned}
(97)\ [Q_{UV}{}^t]^{-1} - [Q_P{}^t]^{-1} &= [1_N \cdot q^1/1_N \cdot q^t] - [p^t \cdot q^1/p^t \cdot q^t] \\
&= \Sigma_{n=1}{}^N [(q_{1n}/1_N \cdot q^t) - (p_{tn} q_{1n}/p^t \cdot q^t)] \\
&= \Sigma_{n=1}{}^N [(1/1_N \cdot q^t) - (p_{tn}/p^t \cdot q^t)] q_{1n} \\
&= \Sigma_{n=1}{}^N [(p^t \cdot q^t/1_N \cdot q^t) - p_{tn}][q_{1n}/p^t \cdot q^t] \\
&= \Sigma_{n=1}{}^N [p_{UV}{}^t - p_{tn}][q_{1n}/p^t \cdot q^t] && \text{using (80) for t = t} \\
&= \Sigma_{n=1}{}^N [p_{UV}{}^t - p_{tn}][q_{1n} - (q_{tn}/Q_{UV}{}^t)]/p^t \cdot q^t && \text{using (87) for t = t} \\
&= [Q_{UV}{}^t]^{-1}\, \Sigma_{n=1}{}^N [p_{UV}{}^t - p_{tn}][(q_{1n} Q_{UV}{}^t) - q_{tn}]/p^t \cdot q^t \\
&= [Q_{UV}{}^t]^{-1}\, \Sigma_{n=1}{}^N s_{tn}[(p_{UV}{}^t/p_{tn}) - 1][(q_{1n} Q_{UV}{}^t/q_{tn}) - 1] && \text{if } q_{tn} > 0 \text{ for all n} \\
&= [Q_{UV}{}^t]^{-1}\, \varepsilon_P{}^t
\end{aligned}
$$

where the period t error term $\varepsilon_P{}^t$ is defined as follows for t = 1,...,T:

$$(98)\ \varepsilon_P{}^t \equiv \Sigma_{n=1}{}^N [p_{UV}{}^t - p_{tn}][(q_{1n} Q_{UV}{}^t) - q_{tn}]/p^t \cdot q^t.$$[85]

[85] Note that this error term is homogeneous of degree 0 in the components of p^t, q^1 and q^t. Thus for $\lambda > 0$, we have $\varepsilon_P(p^t,q^1,q^t) = \varepsilon_P(\lambda p^t,q^1,q^t) = \varepsilon_P(p^t,\lambda q^1,q^t) = \varepsilon_P(p^t,q^1,\lambda q^t)$. Note also that $\varepsilon_P{}^t$ is well defined if some quantities are equal to 0 and $\varepsilon_P{}^t$ does depend on the reservation prices p_{tn} for products n that are not present in period t. If product n is missing in period t, then it is likely that the reservation price p_{tn} is greater than the unit value price level for period t, $p_{UV}{}^t$, and since $q_{tn} = 0$, it can be seen that the nth term on the right hand side of (98) will be negative; i.e., the greater the number of missing products in period t, the greater is the likelihood that $\varepsilon_P{}^t$ is negative.

If $q_{tn} > 0$ for n = 1,...,N, then $\varepsilon_P{}^t$ is equal to $\Sigma_{n=1}{}^N s_{tn}[(p_{UV}{}^t/p_{tn}) - 1][(q_{1n}Q_{UV}{}^t/q_{tn}) - 1]$.

Note that the terms on the right hand side of (97) can be interpreted as $(N/p^t\cdot q^t)$ times the covariance $Cov(p_{UV}{}^t1_N - p^t, q^1 - [Q_{UV}{}^t]^{-1}q^t)$ since $1_N\cdot(q^1 - [Q_{UV}{}^t]^{-1}q^t) = 0$. If the products are substitutable, it is likely that this covariance is *negative*, since if p_{tn} is unusually low, we would expect that it would be less than the period t unit value price $p_{UV}{}^t$ so that $p_{UV}{}^t - p_{tn} > 0$. If p_{tn} is unusually low, then we also expect that the corresponding q_{tn} is unusually high, and thus it is likely that q_{tn} is greater than $q_{1n}Q_{UV}{}^t$ and so $q_{1n}Q_{UV}{}^t - q_{tn} < 0$. Thus the N terms in the covariance will tend to be negative. Thus our expectation is that the error term $\varepsilon_P{}^t < 0$ and $[Q_{UV}{}^t]^{-1} < [Q_P{}^t]^{-1}$ or $Q_{UV}{}^t > Q_P{}^t$ for $t \geq 2$.[86]

There are 3 special cases of (97) which will imply that $Q_{UV}{}^t = Q_P{}^t$: (i) $p^t = \lambda_t 1_N$ so that all period t prices are equal; (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time; (iii) $Cov(p_{UV}{}^t1_N - p^t, q^1 - [Q_{UV}{}^t]^{-1}q^t) = 0$. These conditions are equivalent to our earlier conditions listed below (96).

If we divide both sides of equation t in equations (97) by $[Q_{UV}{}^t]^{-1}$, we obtain the following system of identities for t = 1,...,T:

(99) $Q_P{}^t/Q_{UV}{}^t = [1 - \varepsilon_P{}^t]^{-1}$

where we expect $\varepsilon_P{}^t$ to be a small negative number if the products are substitutable. Thus we expect $Q_P{}^t < Q_{UV}{}^t < Q_L{}^t$ for t = 2,3,...,T.

Equations (97) and (99) are valid if we interchange prices and quantities. Using the definitions for the Dutot price and quantity levels and indexes t and interchanging prices and quantities, equations (99) become $P_P{}^t/P_D{}^t = [1 - \varepsilon_P{}^{t*}]^{-1}$ where $\varepsilon_P{}^{t*} \equiv \Sigma_{n=1}{}^N [q_D{}^t - q_{tn}][(p_{1n}P_D{}^t) - p_{tn}]/p^t\cdot q^t$ for t = 1,...,T. If p_{tn} is unusually low, then it is likely that it will be less than $p_{tn}/P_D{}^t$ and it is also likely that q_{tn} will be unusually high and hence greater than the average period t Dutot quantity level $q_D{}^t$. Thus the N terms in the definition of $\varepsilon_P{}^{t*}$ will tend to be negative and hence a tendency for $[1 - \varepsilon_P{}^{t*}]^{-1}$ to be less than 1. Thus there will be a tendency for $P_P{}^t < P_D{}^t$ for $t \geq 2$.

Equations (94) imply the following identities for t = 1,...,T:

$$(100)\ P_{UV}{}^t/P_L{}^t = Q_P{}^t/Q_{UV}{}^t = [1 - \varepsilon_P{}^t]^{-1}$$

where the last set of equations follow from equations (99). Thus we expect that $P_P{}^t < P_{UV}{}^t < P_L{}^t$ for t = 2,3,...,T if the products are substitutes.

[86] Our expectation that $\varepsilon_P{}^t$ is negative is more strongly held than our expectation that $\varepsilon_L{}^t$ is negative.

Equations (95) and (100) develop exact relationships for the unit value price index P_{UV}^t with the corresponding fixed base Laspeyres and Paasche price indexes, P_L^t and P_P^t. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher price index, P_F^t, and its unit value counterpart period t index, P_{UV}^t, for t = 1,...,T:

(101) $P_{UV}^t = P_F^t\{(1 - \varepsilon_L^t)/(1 - \varepsilon_P^t)\}^{1/2}$

where ε_L^t and ε_P^t are defined by (90) and (98). If there are no strong (divergent) trends in prices and quantities, then it is likely that ε_L^t is approximately equal to ε_P^t and hence under these conditions, it is likely that $P_{UV}^t \approx P_F^t$; i.e., the unit value price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, with diverging trends in prices and quantities (in opposite directions), we would expect the error term ε_P^t defined by (98) to be more negative than the error term ε_L^t defined by (90) and thus under these conditions, we expect the unit value price index P_{UV}^t to have a *downward bias* relative to its Fisher price index counterpart P_F^t.[87]

However, for our empirical example, the unit value price indexes did not approximate their Fisher counterparts very well: the sample averages of the P_{UV}^t and P_F^t were 1.0126 and 0.9745 respectively. The sample averages of the ε_L^t and ε_P^t were –0.1147 and –0.0316 respectively so on average, these error terms were negative as anticipated but the average magnitude of the ε_L^t was very much larger than the average magnitude of the ε_P^t. The reason for the large differences in the averages of the ε_L^t and ε_P^t can be traced to the fact that products 2 and 4 had 0 quantities for observations 1-8. An examination of formula (90) for ε_L^t shows that if some q_{1n} are equal to 0 and the corresponding imputed prices p_{1n} are greater than the unit value price for observation 1, p_{UV}^1, then the nth term in the sum of terms on the right hand side of can become negative and large in magnitude, which can explain why the average magnitude of the ε_L^t was very much larger than the average magnitude of the ε_P^t. Thus we dropped the data for year 1 and started our indexes at the beginning of year 2. The resulting computations showed that the unit value price indexes approximated their Fisher counterparts much more closely: the new sample averages for the P_{UV}^t and P_F^t were 0.7963 and 0.8309 which is in line with our a priori expectations (that the unit value price index was likely to have a downward bias relative to its fixed base Fisher index counterpart). The new sample averages of the ε_L^t and ε_P^t were 0.0046 and – 0.0873 respectively.

It is possible that unit value price indexes can approximate their Fisher counterparts to some degree in some circumstances but these approximations are not likely to be very

[87] The Dutot price index counterparts to the exact relations (101) are $P_F^t = P_D^t\{(1 - \varepsilon_L^{t*})/(1 - \varepsilon_P^{t*})\}^{1/2}$ for t =1,...,T. Thus with diverging trends in prices and quantities (in opposite directions), we would expect the error term ε_P^{t*} to be more negative than the error term ε_L^{t*} and hence we would expect $P_D^t > P_F^t$ for $t \geq 2$. Note that the Dutot price index can be interpreted as a *fixed basket price index* where the basket is proportional to a vector of ones. Thus with divergent trends in prices and quantities in opposite directions, we would expect the Dutot index to exhibit substitution bias and hence we would expect $P_D^t > P_F^t$ for $t \geq 2$.

accurate. If the products are somewhat heterogeneous and there are some divergent trends in price and quantities, then the approximations are likely to be poor.[88] They are also likely to be poor if there is substantial product turnover.

10. Quality Adjusted Unit Value Price and Quantity Indexes

In the previous section, the period t unit value quantity *level* was defined by $q_{UV}^t \equiv 1_N \cdot q^t = \Sigma_{n=1}^N q_{tn}$ for t = 1,...,T. The corresponding period t unit value quantity *index* was defined by (82) for t = 1,...,T: $Q_{UV}^t \equiv 1_N \cdot q^t / 1_N \cdot q^1$. In the present section, we will consider *quality adjusted unit value quantity levels*, $q_{UV\alpha}^t$, and the corresponding quality adjusted *unit value quantity indexes*, $Q_{UV\alpha}^t$, defined as follows for t = 1,...,T:

(102) $q_{UV\alpha}^t \equiv \alpha \cdot q^t$;
(103) $Q_{UV\alpha}^t \equiv q_{UV\alpha}^t / q_{UV\alpha}^1 = \alpha \cdot q^t / \alpha \cdot q^1$

where $\alpha \equiv [\alpha_1,...,\alpha_N]$ is a vector of positive *quality adjustment factors*. Note that if consumers value their purchases of the N products according to the linear utility function $f(q) \equiv \alpha \cdot q$, then the period t quality adjusted aggregate quantity level $q_{UV\alpha}^t = \alpha \cdot q^t$ can be interpreted as the aggregate (sub) *utility* of consumers of the N products. Note that this utility function is linear and thus the products are perfect substitutes, after adjusting for the relative quality of the products. The bigger α_n is, the more consumers will value a unit of product n over other products. The period t *quality adjusted unit value price level* and *price index*, $p_{UV\alpha}^t$ and $P_{UV\alpha}^t$, are defined as follows for t = 1,...,T:

(104) $p_{UV\alpha}^t \equiv p^t \cdot q^t / q_{UV\alpha}^t = p^t \cdot q^t / \alpha \cdot q^t$;
(105) $P_{UV\alpha}^t \equiv p_{UV\alpha}^t / p_{UV\alpha}^1 = [p^t \cdot q^t / p^1 \cdot q^1] / Q_{UV}^t$.

It is easy to check that the quality adjusted unit value price index satisfies Walsh's multiperiod identity test and thus is free from chain drift.[89] Note that the $P_{UV\alpha}^t$ and $Q_{UV\alpha}^t$ *do not depend on the estimated reservation prices*; i.e., the definitions of $P_{UV\alpha}^t$ and $Q_{UV\alpha}^t$ zero out any reservation prices that are applied to missing products. Thus a quality adjusted unit value price index in general *cannot be consistent* with the (Hicksian) economic approach to index number theory if there are new or disappearing products in the sample of products under consideration.

We will start out by comparing $Q_{UV\alpha}^t$ to the corresponding Laspeyres, Paasche and Fisher period t quantity indexes, Q_L^t, Q_P^t and Q_F^t. The algebra in this section follows the algebra

---

[88] The problem with unit value price indexes is that they correspond to an additive quantity level. If one takes the economic approach to index number theory, then an additive quantity level corresponds to a linear utility function which implies an infinite elasticity of substitution between products, which is too high in general.
[89] The term "quality adjusted unit value price index" was introduced by Dalén (2001). Its properties were further studied by de Haan (2004) and de Haan and Krsinich (2018). Von Auer (2014) considered a wide variety of choices for the weight vector α (including $\alpha = p^1$ and $\alpha = p^t$) and he looked at the axiomatic properties of the resulting indexes.

in the preceding section. Thus the counterparts to the identities (87) in the previous section are the following identities for t = 1,...,T:

$$(106)\ \Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}^{t} - p_{tn}]q_{tn} = \Sigma_{n=1}^{N} [\alpha_n(p^t\cdot q^t/\alpha\cdot q^t) - p_{tn}]q_{tn} \quad \text{using definitions (104)}$$
$$= (p^t\cdot q^t/\alpha\cdot q^t)\alpha\cdot q^t - p^t\cdot q^t$$
$$= 0.$$

The difference between the quality adjusted unit value quantity index for period t, $Q_{UV\alpha}^{t}$, and the Laspeyres quantity index for period t, Q_L^t, can be written as follows for t = 1,...,T:

$$(107)\ Q_{UV\alpha}^{t} - Q_L^t = [\alpha\cdot q^t/\alpha\cdot q^1] - [p^1\cdot q^t/p^1\cdot q^1] \quad \text{using (83) and (103)}$$
$$= \Sigma_{n=1}^{N} [(\alpha_n q_{tn}/\alpha\cdot q^1) - (p_{1n}q_{tn}/p^1\cdot q^1)]$$
$$= \Sigma_{n=1}^{N} [(\alpha_n/\alpha\cdot q^1) - (p_{1n}/p^1\cdot q^1)]q_{tn}$$
$$= \Sigma_{n=1}^{N} [(\alpha_n p^1\cdot q^1/\alpha\cdot q^1) - p_{1n}][q_{tn}/p^1\cdot q^1]$$
$$= \Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}^{1} - p_{1n}][q_{tn}/p^1\cdot q^1] \quad \text{using (104) for t = 1}$$
$$= \Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}^{1} - p_{1n}][q_{tn} - q_{1n}Q_{UV\alpha}^{t}]/p^1\cdot q^1 \quad \text{using (106) for t = 1}$$
$$= Q_{UV\alpha}^{t}\ \Sigma_{n=1}^{N} \alpha_n[p_{UV\alpha}^{1} - (p_{1n}/\alpha_n)][(q_{tn}/Q_{UV\alpha}^{t}) - q_{1n}]/p^1\cdot q^1$$
$$= Q_{UV\alpha}^{t}\ \varepsilon_{L\alpha}^{t}$$

where the period t error term $\varepsilon_{L\alpha}^{t}$ is defined for t = 1,...,T as:

$$(108)\ \varepsilon_{L\alpha}^{t} \equiv \Sigma_{n=1}^{N} \alpha_n[p_{UV\alpha}^{1} - (p_{1n}/\alpha_n)][(q_{tn}/Q_{UV\alpha}^{t}) - q_{1n}]/p^1\cdot q^1.$$ [90]

Assuming that $\alpha_n > 0$ for n = 1,...,N, the vector of period t *quality adjusted prices* p_α^t is defined as follows for t = 1,...,T:

$$(109)\ p_\alpha^t \equiv [p_{t1\alpha},..., p_{tN\alpha}] \equiv [p_{t1}/\alpha_1,p_{t2}/\alpha_2,...,p_{tN}/\alpha_N].$$

It can be seen that $p_{UV\alpha}^{1} - (p_{1n}/\alpha_n)$ is the difference between the period 1 unit value price level, $p_{UV\alpha}^{1}$, and the period 1 quality adjusted price for product n, p_{1n}/α_n. Define the period t *quality adjusted quantity share for product n* (using the vector α of quality adjustment factors) as follows for t = 1,...,T and n = 1,...,N:

$$(110)\ S_{tn\alpha} \equiv \alpha_n q_{tn}/\alpha\cdot q^t.$$

The vector of *period t quality adjusted real product shares* (using the vector α of quality adjustment factors) is defined as $S_\alpha^t \equiv [S_{t1\alpha},S_{t2\alpha},...,S_{tN\alpha}]$ for t = 1,...,T. It can be seen that these vectors are share vectors in that their components sum to 1; i.e., we have for t = 1,...,T:

---

[90] This error term is homogeneous of degree 0 in the components of p^1, q^1 and q^t. Hence it is invariant to proportional changes in the components of these vectors. Definition (108) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon_{L\alpha}^{t}$ as $\Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}^{1} - p_{1n}][q_{tn} - q_{1n}Q_{UV\alpha}^{t}]/p^1\cdot q^1$ and with this change, the decomposition defined by the last line of (107) will continue to hold. It should be noted that $\varepsilon_{L\alpha}^{t}$ does not have an interpretation as a *covariance* between a vector of price differences and a vector of quantity differences.

(111) $1_N \cdot S_\alpha^t = 1$.

Using the above definitions, we can show that the period t quality adjusted unit value price level, $p_{UV\alpha}^t$ defined by (104) is equal to a share weighted average of the period t quality adjusted prices $p_{tn\alpha} = p_{tn}/\alpha_n$ defined by (109); i.e., for t = 1,...,T, we have the following equations:

$$\begin{aligned}(112)\ p_{UV\alpha}^t &= p^t \cdot q^t/\alpha \cdot q^t && \text{using (104)} \\ &= \Sigma_{n=1}^N (p_{tn}/\alpha_n)(\alpha_n q_{tn})/\alpha \cdot q^t && \\ &= \Sigma_{n=1}^N S_{tn\alpha} p_{tn\alpha} && \text{using (109) and (110)} \\ &= S_\alpha^t \cdot p_\alpha^t. &&\end{aligned}$$

Now we are in a position to determine the likely sign of $\varepsilon_{L\alpha}^t$ defined by (108). If the products are substitutable, it is likely that $\varepsilon_{L\alpha}^t$ is *negative*, since if p_{1n} is unusually low, then it is likely that the quality adjusted price for product n, p_{1n}/α_n, is below the weighted average of the quality adjusted prices for period 1 which is $p_{UV\alpha}^1 = S_\alpha^1 \cdot p_\alpha^1$ using (112) for t = 1. Thus we expect that $p_{UV\alpha}^1 - (p_{1n}/\alpha_n) > 0$. If p_{1n} is unusually low, then we would expect that the corresponding q_{1n} is unusually high, and thus it is likely that q_{1n} is greater than $q_{tn}/Q_{UV\alpha}^t$ and so $q_{tn}/Q_{UV\alpha}^t - q_{1n} < 0$. Thus the sum of the N terms on the right hand side of (108) is likely to be negative. Thus our expectation[91] is that the error term $\varepsilon_{L\alpha}^t < 0$ and hence $Q_{UV\alpha}^t < Q_L^t$ for $t \geq 2$.

As usual, there are 3 special cases of (108) which will imply that $Q_{UV}^t = Q_L^t$: (i) $p_\alpha^1 = \lambda_1 1_N$ so that all period 1 quality adjusted prices are equal; (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time; (iii) the following sum of price differences times quantity differences equals 0; i.e., $\Sigma_{n=1}^N [\alpha_n p_{UV\alpha}^1 - p_{1n}][(q_{tn}/Q_{UV\alpha}^t) - q_{1n}] = 0$.

If we divide both sides of equation t in equations (108) by Q_{UV}^t, we obtain the following system of identities for t = 1,...,T:

$$(113)\ Q_L^t/Q_{UV\alpha}^t = 1 - \varepsilon_{L\alpha}^t$$

where we expect $\varepsilon_{L\alpha}^t$ to be a small negative number if the products are substitutes.

The difference between reciprocal of the quality adjusted unit value quantity index for period t, $[Q_{UV\alpha}^t]^{-1}$ and the reciprocal of the Paasche quantity index for period t, $[Q_P^t]^{-1}$, can be written as follows for t = 1,...,T:

$$\begin{aligned}(114)\ [Q_{UV\alpha}^t]^{-1} - [Q_P^t]^{-1} &= [\alpha \cdot q^1/\alpha \cdot q^t] - [p^t \cdot q^1/p^t \cdot q^t] && \text{using (84) and (103)} \\ &= \Sigma_{n=1}^N [(\alpha_n q_{1n}/\alpha \cdot q^t) - (p_{tn} q_{1n}/p^t \cdot q^t)] &&\end{aligned}$$

[91] As in the previous section, this expectation is not held with great conviction if the period t quantities have a large variance.

$$
\begin{aligned}
&= \Sigma_{n=1}^{N} [(\alpha_n/\alpha\cdot q^t) - (p_{tn}/p^t\cdot q^t)]q_{1n} \\
&= \Sigma_{n=1}^{N} [(\alpha_n p^t\cdot q^t/\alpha\cdot q^t) - p_{tn}][q_{1n}/p^t\cdot q^t] \\
&= \Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}{}^t - p_{tn}][q_{1n}/p^t\cdot q^t] && \text{using (104)} \\
&= \Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}{}^t - p_{tn}][q_{1n} - (q_{tn}/Q_{UV\alpha}{}^t)]/p^t\cdot q^t && \text{using (106)} \\
&= [Q_{UV\alpha}{}^t]^{-1} \Sigma_{n=1}^{N} \alpha_n[p_{UV\alpha}{}^t - (p_{tn}/\alpha_n)][(q_{1n}Q_{UV\alpha}{}^t) - q_{tn}]/p^t\cdot q^t \\
&= [Q_{UV\alpha}{}^t]^{-1} \varepsilon_{P\alpha}{}^t
\end{aligned}
$$

where the period t error term $\varepsilon_{P\alpha}{}^t$ is defined for t = 1,...,T as:

(115) $\varepsilon_{P\alpha}{}^t \equiv \Sigma_{n=1}^{N} \alpha_n[p_{UV\alpha}{}^t - (p_{tn}/\alpha_n)][(q_{1n}Q_{UV\alpha}{}^t) - q_{tn}]/p^t\cdot q^t$.[92]

If the products are substitutable, it is likely that $\varepsilon_{P\alpha}{}^t$ is *negative*, since if p_{tn} is unusually low, then it is likely that the period t quality adjusted price for product n, p_{tn}/α_n, is below the weighted average of the quality adjusted prices for period t which is $p_{UV\alpha}{}^t = S_\alpha{}^t\cdot p_\alpha{}^t$ using (112). Thus we expect that $p_{UV\alpha}{}^1 - (p_{1n}/\alpha_n) > 0$. If p_{tn} is unusually low, then we would expect that the corresponding q_{tn} is unusually high, and thus it is likely that q_{tn} is greater than $q_{1n}Q_{UV\alpha}{}^t$ and so $q_{1n}Q_{UV\alpha}{}^t - q_{tn} < 0$. Thus the sum of the N terms on the right hand side of (115) is likely to be negative. Thus our expectation is that the error term $\varepsilon_{P\alpha}{}^t < 0$ and hence $[Q_{UV\alpha}{}^t]^{-1} < [Q_L{}^t]^{-1}$ for $t \geq 2$. Assuming that $\varepsilon_{L\alpha}{}^t$ is also negative, we have $Q_P{}^t < Q_{UV\alpha}{}^t < Q_L{}^t$ for t = 2,...,T as inequalities that are likely to hold.

As usual, there are 3 special cases of (114) which will imply that $Q_{UV\alpha}{}^t = Q_P{}^t$: (i) $p_\alpha{}^t = \lambda_t 1_N$ so that all period t quality adjusted prices are equal; (ii) $q^t = \lambda_t q^1$ for t = 2,3,...,T so that quantities vary in strict proportion over time; (iii) the following sum of price differences times quantity differences equals zero: i.e., $\Sigma_{n=1}^{N} [\alpha_n p_{UV\alpha}{}^t - p_{tn}][(q_{1n}Q_{UV\alpha}{}^t) - q_{tn}] = 0$.

If we divide both sides of equation t in equations (114) by $[Q_{UV}{}^t]^{-1}$, we obtain the following system of identities for t = 1,...,T:

(116) $Q_P{}^t/Q_{UV\alpha}{}^t = [1 - \varepsilon_{P\alpha}{}^t]^{-1}$

where we expect $\varepsilon_{P\alpha}{}^t$ to be a small negative number if the products are substitutes.

Equations (113) and (116) develop exact relationships for the quality adjusted unit value quantity index $Q_{UV\alpha}{}^t$ with the corresponding fixed base Laspeyres and Paasche quantity indexes, $Q_L{}^t$ and $Q_P{}^t$. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher quantity index, $Q_F{}^t$, and its quality adjusted unit value counterpart period t quantity index, $Q_{UV\alpha}{}^t$, for t = 1,...,T:

[92] This error term is homogeneous of degree 0 in the components of p^t, q^1 and q^t. Hence it is invariant to proportional changes in the components of these vectors. Definition (115) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon_{P\alpha}{}^t$ as $\Sigma_{n=1}^{N} [\alpha p_{UV\alpha}{}^t - p_{tn}][(q_{1n}Q_{UV\alpha}{}^t) - q_{tn}]/p^t\cdot q^t$ and with this change, the decomposition defined by the last line of (114) will continue to hold..

(117) $Q_F^t = Q_{UV\alpha}^t\{(1 - \varepsilon_{L\alpha}^t)/(1 - \varepsilon_{P\alpha}^t)\}^{1/2}$

where $\varepsilon_{L\alpha}^t$ and $\varepsilon_{P\alpha}^t$ are defined by (108) and (115). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon_{L\alpha}^t$ is approximately equal to $\varepsilon_{P\alpha}^t$ and hence under these conditions, it is likely that $Q_{UV\alpha}^t \approx Q_F^t$; i.e., the quality adjusted unit value quantity index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices in quantities (in opposite directions), then it is likely that $\varepsilon_{P\alpha}^t$ will be more negative than $\varepsilon_{L\alpha}^t$ and hence it is likely that $Q_F^t < Q_{UV\alpha}^t$ for t = 2,...,T; i.e., *with divergent trends in prices and quantities, the quality adjusted unit value quantity index is likely to have an upward bias* relative to its Fisher quantity index counterparts.[93]

Using equations (105), we have the following counterparts to equations (94) for t = 1,...,T:

(118) $p^t \cdot q^t/p^1 \cdot q^1 = P_{UV\alpha}^t Q_{UV\alpha}^t = P_P^t Q_L^t = P_L^t Q_P^t$.

Equations (113), (116) and (118) imply the following identities for t = 1,...,T:

(119) $P_{UV\alpha}^t/P_P^t = Q_L^t/Q_{UV\alpha}^t = 1 - \varepsilon_{L\alpha}^t$;
(120) $P_{UV\alpha}^t/P_L^t = Q_P^t/Q_{UV\alpha}^t = [1 - \varepsilon_{P\alpha}^t]^{-1}$.

We expect that ε_L^t and $\varepsilon_{P\alpha}^t$ will be predominantly negative if the products are substitutes and thus in this case, the quality adjusted unit value indexes $P_{UV\alpha}^t$ should satisfy the inequalities $P_P^t < P_{UV\alpha}^t < P_L^t$ for t = 2,3,...,T.

Taking the square root of the product of equations (119) and (120) leads to the following exact relationships between the fixed base Fisher price index, P_F^t, and its quality adjusted unit value counterpart period t index, P_{UV}^t, for t = 1,...,T:

(121) $P_{UV\alpha}^t = P_F^t\{(1 - \varepsilon_{L\alpha}^t)/(1 - \varepsilon_{P\alpha}^t)\}^{1/2}$

where $\varepsilon_{L\alpha}^t$ and $\varepsilon_{P\alpha}^t$ are defined by (108) and (115). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon_{L\alpha}^t$ is approximately equal to $\varepsilon_{P\alpha}^t$ and hence under these conditions, it is likely that $P_{UV\alpha}^t \approx P_F^t$; i.e., the quality adjusted unit value price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices and quantities, then we expect $\varepsilon_{P\alpha}^t$ to be more negative than $\varepsilon_{L\alpha}^t$ and hence there is an expectation that $P_{UV\alpha}^t < P_F^t$ for t = 2,...,T; i.e., we expect that normally $P_{UV\alpha}^t$ will have a

[93] As was the case in the previous section, if there are missing products in period 1, the expected inequality $P_{UV\alpha}^t < P_F^t$ may be reversed, because $\varepsilon_{L\alpha}^t$ defined by (108) may become significantly negative if some q_{1n} = 0 while their corresponding reservation prices $p_{1n} > 0$.

downward bias relative to P_F^t.[94] However, if there are missing products in period 1, then the bias of $P_{UV\alpha}^t$ relative to P_F^t is uncertain.

11. Relationships between Lowe and Fisher Indexes

We now consider how a Lowe (1823) price index is related to a fixed base Fisher price index. The framework that we consider is similar to the framework developed in section 6 above for the annual share weighted Jevons index, $P_{J\alpha}^t$. In the present section, instead of using the average sales shares for the first year in the sample as weights for a weighted Jevons index, we use annual average quantities sold in the first year as a vector of quantity weights for subsequent periods. Define the *annual average quantity vector* $q^* \equiv [q_1^*,...,q_N^*]$ for the first T^* periods in the sample that make up a year, q^*, as follows:[95]

(122) $q^* \equiv (1/T^*)\Sigma_{t=1}^{T^*} q^t$.

As was the case in section 6, the reference year for the weights precedes the reference month for the product prices. Define the *period t Lowe* (1823) *price level* and *price index*, p_{Lo}^t and P_{Lo}^t by (123) and (124) respectively for $t = T^*+1,T^*+2,...,T$:

(123) $p_{Lo}^t \equiv p^t \cdot \alpha$;
(124) $P_{Lo}^t \equiv p_{Lo}^t/p_{Lo}^{T^*+1} = p^t \cdot \alpha/p^{T^*+1} \cdot \alpha$

where the constant price weights vector α is defined as the annual average weights vector q^* defined by (122); i.e., we have:

(125) $\alpha \equiv q^*$.

The *period t Lowe quantity level*, q_{Lo}^t, and the corresponding *period t Lowe quantity index*, Q_{Lo}^t, are defined as follows for $t = T^*+1,T^*+2,...,T$:

(126) $q_{Lo}^t \equiv p^t \cdot q^t/p_{Lo}^t = p^t \cdot q^t/p^t \cdot \alpha = \Sigma_{n=1}^N (p_{tn}\alpha_n)(q_{tn}/\alpha_n)/p^t \cdot \alpha$ [96]
(127) $Q_{Lo}^t \equiv q_{Lo}^t/q_{Lo}^{T^*+1} = [p^t \cdot q^t/p^{T^*+1} \cdot q^{T^*+1}]/P_{Lo}^t$.

[94] Recall that the weighted unit value quantity level, $q_{UV\alpha}^t$ is defined as the linear function of the period t quantity data, $\alpha \cdot q^t$. If $T \geq 3$ and the price and quantity data are consistent with purchasers maximizing a utility function that generates data that is exact for the Fisher price index Q_F^t, then $Q_{UV\alpha}^t$ will tend to be greater than Q_F^t (and hence $P_{UV\alpha}^t$ will tend to be less than P_F^t) for $t \geq 2$. See Marris (1984; 52), Diewert (1999; 49) and Diewert and Fox (2017; 26) on this point.

[95] If product n was not available in the first year of the sample, then the nth component of q^*, q_n^*, will equal 0 and hence the nth component of the weight vector α defined by (125) will also equal 0. If product n was also not available in periods $t \geq T^* + 1$, then looking at definitions (123) and (124), it can be seen that P_{LO}^t will not depend on the reservation prices p_{nt} for these subsequent periods where product n is not available. Thus under these circumstances, the Lowe index cannot be consistent with the (Hicksian) economic approach to index number theory since Konüs (1924) true cost of living price indexes will depend on the reservation prices.

[96] This last inequality is only valid if all $\alpha_n > 0$.

It can be seen that the Lowe price index defined by (124) is equal to a *weighted Dutot price index*; see definition (14) above. It is also structurally identical to the quality adjusted unit value quantity index $Q_{UV\alpha}{}^t$ defined in the previous section, except the role of prices and quantities has been reversed. Thus the identity (107) in the previous section will be valid if we replace $Q_{UV\alpha}{}^t$ by $P_{Lo}{}^t$, replace $Q_L{}^t$ by $P_L{}^t$ and interchange prices and quantities on the right hand side of (107).[97] The resulting identities are the following ones for $t = T^*+1,T^*+2,...,T$:

$$
\begin{aligned}
(128)\quad P_{Lo}{}^t - P_L{}^t &= \Sigma_{n=1}{}^N\,[(\alpha_n\, p_{tn}/\alpha\cdot p^{T^*+1}) - (p_{tn}q_{T^*+1,n}/p^{T^*+1}\cdot q^{T^*+1})] \\
&= \Sigma_{n=1}{}^N\,[(\alpha_n/\alpha\cdot p^{T^*+1}) - (q_{T^*+1,n}/p^{T^*+1}\cdot q^{T^*+1})]p_{tn} \\
&= \Sigma_{n=1}{}^N\,[(\alpha_n p^{T^*+1}\cdot q^{T^*+1}/\alpha\cdot p^{T^*+1}) - q_{T^*+1,n}][p_{tn}/p^{T^*+1}\cdot q^{T^*+1}] \\
&= \Sigma_{n=1}{}^N\,[\alpha_n q_{Lo}{}^{T^*+1} - q_{T^*+1,n}][p_{tn}/p^{T^*+1}\cdot q^{T^*+1}] \qquad \text{using (126) for } t = T^*+1 \\
&= \Sigma_{n=1}{}^N\,[\alpha_n q_{Lo}{}^{T^*+1} - q_{T^*+1,n}][p_{tn} - p_{T^*+1,n}P_{Lo}{}^t]/p^{T^*+1}\cdot q^{T^*+1}\ {}^{98} \\
&= P_{Lo}{}^t\, \Sigma_{n=1}{}^N\,[\alpha_n q_{Lo}{}^{T^*+1} - q_{T^*+1,n}][(p_{tn}/P_{Lo}{}^t) - p_{T^*+1,n}]/p^{T^*+1}\cdot q^{T^*+1} \\
&= P_{Lo}{}^t\, \Sigma_{n=1}{}^N\,\alpha_n[q_{Lo}{}^{T^*+1} - (q_{T^*+1,n}/\alpha_n)][(p_{tn}/P_{Lo}{}^t) - p_{T^*+1,n}]/p^{T^*+1}\cdot q^{T^*+1} \\
&= P_{Lo}{}^t\, \varepsilon_{L\alpha}{}^t
\end{aligned}
$$

where the period t error term $\varepsilon_{L\alpha}{}^t$ is now defined for $t = T^*+1,...,T$ as follows:

$$(129)\quad \varepsilon_{L\alpha}{}^t \equiv \Sigma_{n=1}{}^N\,\alpha_n[q_{Lo}{}^{T^*+1} - (q_{T^*+1,n}/\alpha_n)][(p_{tn}/P_{Lo}{}^t) - p_{T^*+1,n}]/p^{T^*+1}\cdot q^{T^*+1}.{}^{99}$$

If the products are substitutable, it is likely that $\varepsilon_{L\alpha}{}^t$ is *negative*, since if $p_{T^*+1,n}$ is unusually low, then it is likely that $(p_{tn}/P_{Lo}{}^t) - p_{T^*+1,n} > 0$ and that $q_{T^*+1,n}/\alpha_n$ is unusually large and hence is greater than $q_{Lo}{}^t$, which is a weighted average of the period T^*+1 quantity ratios, $q_{T^*+1,1}/\alpha_1$, $q_{T^*+1,2}/\alpha_2$,..., $q_{T^*+1,N}/\alpha_N$ using definition (126) for $t = T^*+1$. Thus the sum of the N terms on the right hand side of (129) is likely to be negative. Thus our expectation[100] is that the error term $\varepsilon_{L\alpha}{}^t < 0$ and hence $P_{Lo}{}^t < P_L{}^t$ for $t > T^* + 1$.

The α_n can be interpreted as *inverse quality indicators* of the utility provided by one unit of the nth product. Suppose purchasers of the N commodities have Leontief preferences with the utility function $f(q_1,q_2,...,q_N) \equiv \min_n \{q_n/\alpha_n : n = 1,2,...,N\}$. Then the dual unit cost function that corresponds to this functional form is $c(p_1,p_2,...,p_N) \equiv \Sigma_{n=1}{}^N\, p_n\alpha_n = p\cdot\alpha$. If we evaluate the unit cost function at the prices of period t, p^t, we obtain the Lowe price level for period t defined by (123); i.e., $p_{Lo}{}^t \equiv p^t\cdot\alpha$. Thus the bigger α_n is, the more units of q_n it will take for purchasers of the N commodities to attain one unit of utility. Thus the α_n can be interpreted as inverse indicators of the relative utility of each product.

---

[97] We also replace period 1 by period T^*+1.

[98] This step follows using the following counterpart to (106): $\Sigma_{n=1}{}^N\,[\alpha_n q_{Lo}{}^{T^*+1} - q_{T^*+1,n}]p_{T^*+1,n} = 0$.

[99] Note that this error term is homogeneous of degree 0 in the components of p^{T^*+1}, q^{T^*+1} and p^t. Hence it is invariant to proportional changes in the components of these vectors. Definition (129) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon_{L\alpha}{}^t$ as $\Sigma_{n=1}{}^N\,[\alpha_n q_{Lo}{}^{T^*+1} - q_{T^*+1,n}][(p_{tn}/P_{Lo}{}^t) - p_{T^*+1,n}]/p^{T^*+1}\cdot q^{T^*+1}$ and with this change, the decomposition defined by the last line of (128) will continue to hold.

[100] This expectation is not held with great conviction if the period t prices have a large variance.

As usual, there are 3 special cases of (128) which will imply that $P_{Lo}^t = P_L^t$: (i) $q^{T*+1} = \lambda q^*$ for some $\lambda > 0$ so that the period T^*+1 quantity vector q^{T*+1} is proportional to the annual average quantity vector q^* for the base year; (ii) $p^t = \lambda_t p^{T*+1}$ for some $\lambda_t > 0$ for $t = T^*+1,...,T$ so that prices vary in strict proportion over time; (iii) the sum of terms $\Sigma_{n=1}^N [\alpha_n q_{Lo}^{T*+1} - q_{T*+1,n}][(p_{tn}/P_{Lo}^t) - p_{T*+1,n}] = 0$.

If we divide both sides of equation t in equations (128) by P_{Lo}^t, we obtain the following system of identities for $t = T^*+1,...,T$:

(130) $P_L^t/P_{Lo}^t = 1 - \varepsilon_{L\alpha}^t$

where we expect $\varepsilon_{L\alpha}^t$ to be a small negative number.

We turn now to developing a relationship between the Lowe and Paasche price indexes. The difference between reciprocal of the Lowe price index for period t, $[P_{Lo}^t]^{-1}$ and the reciprocal of the Paasche price index for period t, $[P_P^t]^{-1}$, can be written as follows for $t = T^*+1,...,T$:

$$
\begin{aligned}
(131)\ [P_{Lo}^t]^{-1} - [P_P^t]^{-1} &= [\alpha\cdot p^{T*+1}/\alpha\cdot p^t] - [q^t\cdot p^{T*+1}/q^t\cdot p^t] \\
&= \Sigma_{n=1}^N [(\alpha_n p_{T*+1,n}/\alpha\cdot p^t) - (q_{tn}p_{T*+1,n}/p^t\cdot q^t)] \\
&= \Sigma_{n=1}^N [(\alpha_n/\alpha\cdot p^t) - (q_{tn}/p^t\cdot q^t)]p_{T*+1,n} \\
&= \Sigma_{n=1}^N [(\alpha_n p^t\cdot q^t/\alpha\cdot q^t) - q_{tn}][p_{T*+1,n}/p^t\cdot q^t] \\
&= \Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}][p_{T*+1,n}/p^t\cdot q^t] && \text{using (126)} \\
&= \Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}][p_{T*+1,n} - (p_{tn}/P_{Lo}^t)]/p^t\cdot q^t\ {}^{101} \\
&= [P_{Lo}^t]^{-1}\, \Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}][p_{T*+1,n}P_{Lo}^t - p_{tn}]/p^t\cdot q^t \\
&= [P_{Lo}^t]^{-1}\, \Sigma_{n=1}^N \alpha_n[q_{Lo}^t - (q_{tn}/\alpha_n)][p_{T*+1,n}P_{Lo}^t - p_{tn}]/p^t\cdot q^t && \text{if all } \alpha_n > 0 \\
&= [P_{Lo}^t]^{-1}\, \varepsilon_{P\alpha}^t
\end{aligned}
$$

where the period t error term $\varepsilon_{P\alpha}^t$ is defined for $t = T^*+1,...,T$ as:

(132) $\varepsilon_{P\alpha}^t \equiv \Sigma_{n=1}^N \alpha_n[q_{Lo}^t - (q_{tn}/\alpha_n)][p_{T*+1,n}P_{Lo}^t - p_{tn}]/p^t\cdot q^t$.[102]

If the products are substitutable, it is likely that $\varepsilon_{P\alpha}^t$ is *negative*, since if p_{tn} is unusually low, then it is likely that it will be less than the inflation adjusted nth component of the period T^*+1 price, $p_{T*+1,n}P_{LO}^t$. If p_{tn} is unusually low, then it is also likely that the period t quality adjusted quantity for product n, q_{tn}/α_n, is above the weighted average of the quality adjusted quantities for period t which is q_{Lo}^t. Thus the sum of the N terms on the right hand side of (132) is likely to be negative. Thus our expectation is that the error term $\varepsilon_{P\alpha}^t < 0$ and hence $[P_{Lo}^t]^{-1} < [P_P^t]^{-1}$ for $t \geq 2$. Assuming that $\varepsilon_{L\alpha}^t$ is also negative, we have $P_P^t < P_{Lo}^t < P_L^t$ for $t = T^*+2,T^*+3,...,T$ as inequalities that are likely to hold.

---

[101] This step follows using the following counterpart to (106): $\Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}]p_{tn} = 0$.

[102] This error term is homogeneous of degree 0 in the components of q^t, p^{T*+1} and p^t. Hence it is invariant to proportional changes in the components of these vectors. Definition (132) is only valid if all $\alpha_n > 0$. If this is not the case, redefine $\varepsilon_{P\alpha}^t$ as $\Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}][p_{T*+1,n}P_{Lo}^t - p_{tn}]/p^t\cdot q^t$ and with this change, the decomposition defined by the last line of (131) will continue to hold.

As usual, there are 3 special cases of (131) which will imply that $P_{Lo}^t = P_P^t$: (i) $q^t = \lambda q^*$ for some $\lambda > 0$ so that the period t quantity vector q^t is proportional to the annual average quantity vector q^* for the reference year prior to the reference month; (ii) $p^t = \lambda_t p^{T^*+1}$ for t $= T^*+2, T^*+3, ..., T$ so that prices vary in strict proportion over time; (iii) the sum of terms $\Sigma_{n=1}^N [\alpha_n q_{Lo}^t - q_{tn}][p_{T^*+1,n} P_{Lo}^t - p_{tn}] = 0$.

If we divide both sides of equation t in equations (131) by $[P_{Lo}^t]^{-1}$, we obtain the following system of identities for t = 1,...,T:

(133) $P_P^t/P_{Lo}^t = [1 - \varepsilon_{P\alpha}^t]^{-1}$

where we expect $\varepsilon_{P\alpha}^t$ to be a negative number.

Equations (130) and (133) develop exact relationships for the Lowe price index P_{Lo}^t with the corresponding fixed base Laspeyres and Paasche price indexes, P_L^t and P_P^t. Taking the square root of the product of these two sets of equations leads to the following exact relationships between the fixed base Fisher price index, P_F^t, and the corresponding Lowe period t price index, P_{Lo}^t, for $t = T^*+1, ..., T$:

(134) $P_F^t = P_{Lo}^t\{(1 - \varepsilon_{L\alpha}^t)/(1 - \varepsilon_{P\alpha}^t)\}^{1/2}$

where $\varepsilon_{L\alpha}^t$ and $\varepsilon_{P\alpha}^t$ are defined by (129) and (132). If there are no strong (divergent) trends in prices and quantities, then it is likely that $\varepsilon_{L\alpha}^t$ is approximately equal to $\varepsilon_{P\alpha}^t$ and hence under these conditions, it is likely that $P_{Lo}^t \approx P_F^t$; i.e., the Lowe price index will provide an adequate approximation to the fixed base Fisher price index under these conditions. However, if there are divergent trends in prices in quantities (in diverging directions), then it is likely that $\varepsilon_{P\alpha}^t$ will be more negative than $\varepsilon_{L\alpha}^t$ and hence it is likely that $P_F^t < P_{Lo}^t$ for $t = T^*+2, ..., T$; i.e., *with divergent trends in prices and quantities, the Lowe price index is likely to have an upward bias* relative to its Fisher Price index counterpart.

In section 6 above, we defined a weighted Jevons index (or Geometric Young index), $P_{J\alpha}^t$, that set the vector of weights α equal to the average expenditure shares in the first year of our sample. For our frozen juice data listed in the Appendix, $P_{J\alpha}^t$ was compared to the geometric Paasche, Laspeyres and Törnqvist price indexes, P_{GP}^t, P_{GL}^t and P_T^t respectively, for the 26 periods running from “month” 14 to 39. The resulting means for the P_{GP}^t, P_T^t, $P_{J\alpha}^t$, P_{GL}^t were 0.81005, 0.83100, 0.83338 and 0.85270 respectively. In the present section, we compare the Lowe price indexes, P_{Lo}^t, to the unit value index P_{UV}^t and to the fixed base Törnqvist, Fisher, Laspeyres and Paasche price indexes, P_T^t, P_F^t, P_L^t and P_P^t respectively for the same 26 periods which follow the first year in our sample. The resulting means for the P_{Lo}^t, P_{UV}^t, P_T^t, P_F^t, P_L^t and P_P^t were 0.83772, 0.79634, 0.83100, 0.83094, 0.86329 and 0.80014 respectively.[103] Thus on average, the Lowe indexes P_{Lo}^t were 0.68 percentage points above the corresponding Fisher indexes P_F^t. This is a substantial upward bias.

In the following section, we show that the Geary Khamis multilateral indexes can be regarded as quality adjusted unit value price indexes and hence the analysis in section 10 on quality adjusted unit value price indexes can be applied to these multilateral indexes.

12. Geary Khamis Multilateral Indexes

The GK multilateral method was introduced by Geary (1958) in the context of making international comparisons of prices. Khamis (1970) showed that the equations that define the method have a positive solution under certain conditions. A modification of this method has been adapted to the time series context and is being used to construct some components of the Dutch CPI; see Chessa (2016).

The GK system of equations for T time periods involves T *price levels* $p_{GK}^1,...,p_{GK}^T$ and N *quality adjustment factors* $\alpha_1,...,\alpha_N$.[104] Let p^t and q^t denote the N dimensional price and quantity vectors for period t (with components p_{tn} and q_{tn} as usual). Define the total consumption vector q over the entire window as the following simple sum of the period by period consumption vectors:

(135) $q \equiv \sum_{t=1}^T q^t$

where $q \equiv [q_1,q_2,...,q_N]$. The equations which determine the *GK price levels* $p_{GK}^1,...,p_{GK}^T$ and *quality adjustment factors* $\alpha_1,...,\alpha_N$ (up to a scalar multiple) are the following ones:

(136) $\alpha_n = \sum_{t=1}^T [q_{tn}/q_n][p_{tn}/p_{GK}^t]$; $\quad n = 1,...,N;$

(137) $p_{GK}^t = p^t \cdot q^t/\alpha \cdot q^t = \Sigma_{n=1}^N [\alpha_n q_{tn}/\alpha \cdot q^t][p_{tn}/\alpha_n]$; $\quad t = 1,...,T$

where $\alpha \equiv [\alpha_1,...,\alpha_N]$ is the vector of GK quality adjustment factors. The sample share of period t's purchases of commodity n in total sales of commodity n over all T periods can be defined as $S_{tn} \equiv q_{tn}/q_n$ for n = 1,...,N and t = 1,...,T. Thus $\alpha_n \equiv \Sigma_{t=1}^T S_{tn}[p_{tn}/p_{GK}^t]$ is a (real) share weighted average of the period t inflation adjusted prices p_{tn}/ p_{GK}^t for product n over all T periods. The period t quality adjusted sum of quantities sold is defined as the *period t GK quantity level*, $q_{GK}^t \equiv \alpha \cdot q^t = \Sigma_{n=1}^N \alpha_n q_{tn}$.[105] This period t quantity level is divided into the value of period t sales, $p^t \cdot q^t = \Sigma_{n=1}^N p_{tn}q_{tn}$, in order to obtain the period t GK price level, p_{GK}^t. Thus the GK price level for period t can be interpreted as a *quality adjusted unit value index* where the α_n act as the quality adjustment factors.

[103] The mean of the error terms ε_L^t was –0.0309 and the mean of the error terms ε_P^t was –0.0488. Thus on average, both error terms were negative and the ε_P^t tended to be more negative than the ε_L^t, which is consistent with our expectations.

[104] In the international context, the α_n are interpreted as international commodity reference prices.

[105] Khamis (1972; 101) also derived this equation in the time series context.

Note that the GK price level, p_{GK}^t defined by (137) *does not depend on the estimated reservation prices*; i.e., the definition of p_{GK}^t zeros out any reservation prices that are applied to missing products and thus P_{GK}^t also does not depend on reservation prices. Thus the GK price indexes in general *cannot be consistent* with the (Hicksian) economic approach to index number theory if there are new or disappearing products in the sample of products under consideration.

It can be seen that if a solution to equations (136) and (137) exists, then if all of the period price levels p_{GK}^t are multiplied by a positive scalar λ say and all of the quality adjustment factors α_n are divided by the same λ, then another solution to (136) and (137) is obtained. Hence, the α_n and p_{GK}^t are only determined up to a scalar multiple and an additional normalization is required such as $p_{GK}^1 = 1$ or $\alpha_1 = 1$ is required to determine a unique solution to the system of equations defined by (136) and (137).[106] It can also be shown that only N + T − 1 of the N + T equations in (136) and (137) are independent.

A traditional method for obtaining a solution to (136) and (137) is to iterate between these equations. Thus set $\alpha = 1_N$, a vector of ones and use equations (137) to obtain an initial sequence for the P_{GK}^t. Substitute these P_{GK}^t estimates into equations (136) and obtain α_n estimates. Substitute these α_n estimates into equations (137) and obtain a new sequence of P_{GK}^t estimates. Continue iterating between the two systems until convergence is achieved.

An alternative method is more efficient. Following Diewert (1999; 26),[107] substitute equations (71) into equations (70) and after some simplification, obtain the following system of equations which will determine the components of the α vector:

$$(138)\ [I_N - C]\alpha = 0_N$$

where I_N is the N by N identity matrix, 0_N is a vector of zeros of dimension N and the C matrix is defined as follows:

$$(139)\ C \equiv \hat{q}^{-1}\,\Sigma_{t=1}^{T} s^t q^{tT} = \Sigma_{t=1}^{T} s^t\ \hat{q}^{-1} q^{tT} = \Sigma_{t=1}^{T} s^t\, S^{tT}$$

where $\hat{q}$ is an N by N diagonal matrix with the elements of the total window purchase vector q running down the main diagonal and $\hat{q}^{-1}$ denotes the inverse of this matrix, s^t is the period t expenditure share column vector, q^t is the column vector of quantities purchased during period t, S^{tT} is the period t real sample product share row vector $[S_{t1},S_{t2},...S_{tN}]$ where $S_{tn} \equiv q_{tn}/q_n$ for n = 1,...,N and t = 1,...,T and q_n is the nth element of the sample q defined by (135).

The matrix $I_N - C$ is singular which implies that the N equations in (138) are not all independent. In particular, if the first N−1 equations in (138) are satisfied, then the last

[106] See Diewert and Fox (2017) for various solution methods.
[107] See also Diewert and Fox (2017; 33) for additional discussion on this solution method.

equation in (138) will also be satisfied. It can also be seen that the N equations in (138) are homogeneous of degree one in the components of the vector α. Thus to obtain a unique b solution to (138), set α_N equal to 1, drop the last equation in (138) and solve the remaining N−1 equations for $\alpha_1,\alpha_2,...,\alpha_{N-1}$. This is the solution method we used in this study.

Using equations (137), it can be seen that the *GK price index for period t* is equal to $P_{GK}{}^t \equiv p_{GK}{}^t/p_{GK}{}^1 = [p^t{\cdot}q^t/\alpha{\cdot}q^t]/[p^1{\cdot}q^1/\alpha{\cdot}q^1]$ for t = 1,...,T and thus these indexes are *quality adjusted unit value price indexes* with a particular choice for the vector of quality adjustment factors α. Thus these indexes lead to corresponding *additive quantity levels* $q_{GK}{}^t$ that correspond to the linear utility function, $f(q) \equiv \alpha{\cdot}q$.[108] As we saw in section 10, this type of index can approximate the corresponding fixed base Fisher price index provided that there are no systematic divergent trends in prices and quantities. However, if there are diverging trends in prices and quantities (in opposite directions), then we expect the GK price indexes to be subject to some *substitution bias* with the expectation that the GK price index for period $t \geq 2$ to be somewhat *below* the corresponding Fisher fixed base price index.

When we calculated the GK price indexes for our empirical example over our 3 years of data, we found that the average value of the GK price indexes was 0.99764 while the corresponding average values for the fixed base Fisher and Törnqvist indexes were 0.97434 and 0.97607 so that on average, $P_{GK}{}^t$ was about 2 percentage points *above* the corresponding $P_F{}^t$ and $P_T{}^t$. This is a substantial *upward* bias instead of the expected *downward* bias. However, our data for the first year are not representative for the remaining two years; i.e., products 2 and 4 were missing for the first 2/3 of the first year and this affects the GK indexes. When we dropped the first year of data from our sample and calculated GK, fixed base Fisher, Törnqvist, Laspeyres, Paasche and unit value price indexes for the remaining two years of our sample, the sample average values for $P_{GK}{}^t$, $P_F{}^t$, $P_T{}^t$, $P_L{}^t$, $P_P{}^t$ and $P_{UV}{}^t$ were 0.81966, 0.83094, 0.83100, 0.86329, 0.80014 and 0.79634 respectively. Thus for the smaller sample, $P_{GK}{}^t$ was about 1 percentage point *below* the corresponding $P_F{}^t$ and $P_T{}^t$, which is in line with our a priori expectations. Note that the average sample value for the unit value price index for the smaller sample was 0.79634, well *below* the corresponding $P_F{}^t$ and $P_T{}^t$ averages, which is also in line with our a priori expectations. Thus we expect GK and quality adjusted unit value price indexes to *normally* have a downward bias relative to their Fisher and Törnqvist counterparts, provided that there are no missing products, the products are highly substitutable and there are divergent trends in prices and quantities.

13. Weighted Time Product Dummy Multilateral Indexes

The time product dummy multilateral indexes are motivated by the following model of price behavior:

[108] Thus the GK price indexes will be exactly the correct price indexes to use if purchasers maximize utility using a common linear utility function. Diewert (1999; 27) and Diewert and Fox (2017; 33-34) show that the GK price indexes will also be exactly correct if purchasers maximize a Leontief, no substitution utility function. These extreme cases are empirically unlikely.

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

The parameter π_t can be interpreted as the *time product dummy price level* for period t, α_n can be interpreted as a commodity n *quality adjustment factor* and e_{tn} is a stochastic error term with mean that is assumed to be 1. Define the logarithms of p_{tn} and e_{tn} as $y_{tn} \equiv \ln p_{tn}$ and $\varepsilon_{tn} \equiv \ln e_{tn}$ for $t = 1,...,T$; $n = 1,...,N$, define the logarithm of π_t as $\rho_t \equiv \ln\pi_t$ for $t = 1,...,T$ and define the logarithm of α_n as $\beta_n \equiv \ln\alpha_n$ for $n = 1,...,N$. Then taking logarithms of both sides of (140) leads to the following linear regression model:

(141) $y_{tn} = \rho_t + \beta_n + \varepsilon_{tn}$; $t = 1,...,T;\ n = 1,...,N.$

The ρ_t and β_n can be estimated by solving a least squares minimization problem.[109] This is Summer's (1973) country product dummy multilateral method adapted to the time series context.

Rao (1995)[110] suggested the following weighted by economic importance version of Summer's method: find the ρ_t and β_n which solve the following *weighted least squares minimization problem*:

(142) $\min_{\rho_1,...,\rho_T,\beta_1,...,\beta_N} \Sigma_{t=1}^{T}\Sigma_{n=1}^{N} s_{tn}(y_{tn} - \rho_t - \beta_n)^2.$

The first order necessary (and sufficient) conditions for solving (142) are the following T equations (143) and the N equations (144):

(143) $\rho_t + \Sigma_{n=1}^{N} s_{tn}\beta_n = \Sigma_{n=1}^{N} s_{tn}y_{tn}$; $t = 1,...,T;$
(144) $\Sigma_{t=1}^{T} s_{tn}\rho_t + (\Sigma_{t=1}^{T} s_{tn})\beta_n = \Sigma_{t=1}^{T} s_{tn}y_{tn}$; $n = 1,...,N.$

Looking at equations (143) and (144), it can be seen that solutions ρ_t and β_n to these equations *do not depend on any reservation prices*; i.e., if y_{tn} is the logarithm of a reservation price, it is always multiplied by the 0 share s_{tn} in the above equations. Thus as was the case with unit value and GK price indexes, any reservation prices are multiplied by 0 shares or quantities, which nullifies their effect on the resulting indexes. Thus if there are missing products in the sample, the weighted time product dummy price indexes cannot be consistent with the Hicksian economic approach to index number theory.

For our empirical example in the Appendix, we used equations (143) and (144) to solve for the T log price levels ρ_t and the N log quality adjustment factors β_n.[111] The resulting

[109] A normalization on the parameters such as $\beta_N = 0$ (which corresponds to $\alpha_N = 1$) is required to identify the parameters.
[110] See also Diewert (2004) (2005) (2012) and Rao (1995) (2005) on the WTPD method. Balk (1980; 70) suggested the use of the weighted time product dummy method to the time series context but used different weights. Ivancic, Diewert and Fox (2009) applied the method using the weights in (142) in the time series context.
[111] Equations (143) and (144) are linearly dependent; see Diewert and Fox (2017). Thus we dropped one of these equations and set $\beta_N = 0$, which means we set $\alpha_N = 1$.

solution for ρ_t is equal to the logarithm of the *period t weighted time product dummy price level*, p_{WTPD}^t, and the solution for β_n is the logarithm of the quality adjustment factor for product n, α_n. Thus we have:

(145) $p_{WTPD}^t \equiv \exp[\rho_t]$; $t = 1,...,T$; $\alpha_n \equiv \exp[\beta_n]$; $n = 1,...,N$.

Using definitions (145) and the definitions $y_{tn} \equiv \ln p_{tn}$, equations (143) can be rewritten as follows:

(146) $\ln p_{WTPD}^t = \Sigma_{n=1}^N s_{tn}(\ln p_{tn} - \beta_n) = \Sigma_{n=1}^N s_{tn}\ln(p_{tn}/\alpha_n)$; $t = 1,...,T$.

Thus the period t *weighted time product dummy price level*, p_{WTPD}^t, is equal to the following *quality adjusted weighted Jevons price level*:

(147) $p_{WTPD}^t = \Pi_{n=1}^N (p_{tn}/\alpha_n)^{s_{tn}}$; $t = 1,...,T$.

Equations (144) and definitions (145) can be used to derive the following expressions for the logarithms of the product quality adjustment parameters:[112]

(148) $\ln\alpha_n \equiv \beta_n = \Sigma_{t=1}^T s_{tn}\ln(p_{tn}/p_{WTPD}^t)/\Sigma_{t=1}^T s_{tn}$; $n = 1,...,N$.

Thus β_n is equal to a share weighted average of the logarithms of the sample prices for product n, deflated by the weighted time product dummy price levels for each period; i.e., β_n is equal to *a weighted average (over t) of the logarithms of the inflation adjusted prices for product n*, p_{tn}/p_{WTPD}^t.

Once the WTPD price levels have been defined, the *weighted time product dummy price index* for period t is defined as $P_{WTPD}^t \equiv p_{WTPD}^t/p_{WTPD}^1$ and the logarithm of P_{WTPD}^t is equal to the following expression:

(149) $\ln P_{WTPD}^t = \Sigma_{n=1}^N s_{tn}(\ln p_{tn} - \beta_n) - \Sigma_{n=1}^N s_{1n}(\ln p_{1n} - \beta_n)$; $t = 1,...,T$.

With the above expression for $\ln P_{WTPD}^t$ in hand, we can compare $\ln P_{WTPD}^t$ to $\ln P_T^t$. Using (149) and definition (40), we can derive the following expressions for $t = 1,2,...,T$:

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

Since $\Sigma_{n=1}^N (s_{tn} - s_{t1}) = 0$ for each t, the two sets of terms on the right hand side of equation t in (150) can be interpreted as normalizations of the covariances between $s^t - s^1$ and $\ln p^t - \beta$ for the first set of terms and between $s^t - s^1$ and $\ln p^1 - \beta$ for the second set of terms. If the products are highly substitutable with each other, then a low p_{tn} will usually

---

[112] If p_{tn} is a reservation price, then the nth term on the right hand side of equation t in (147) becomes $(p_{tn}/\alpha_n)^0 = 1$ and the t^{th} term on the right hand side of equation n in (148) becomes 0. Thus any effect of differing reservation prices on the WTPD price levels and quality adjustment factors is nullified.

imply that $\ln p_{tn}$ is less than the average log price β_n and it is also likely that s_{tn} is greater than s_{1n} so that $(s_{tn} - s_{1n})(\ln p_{tn} - \beta_n)$ is likely to be negative. Hence the covariance between $s^t - s^1$ and $\ln p^t - \beta$ will tend to be negative. On the other hand, if p_{1n} is unusually low, then $\ln p_{tn}$ will be less than the average log price β_n and it is likely that s_{1n} is greater than s_{tn} so that $(s_{tn} - s_{1n})(\ln p_{1n} - \beta_n)$ is likely to be positive. Hence the covariance between $s^t - s^1$ and $\ln p^1 - \beta$ will tend to be positive. Thus the first set of terms on the right hand side of (150) will tend to be negative while the second set will tend to be positive. If there are no divergent trends in log prices and sales shares, then it is likely that these two terms will largely offset each other and under these conditions, $P_{WTPD}{}^t$ is likely to approximate $P_T{}^t$ reasonably well. However, with divergent trends and highly substitutable products, it is likely that the first set of negative terms will be larger in magnitude than the second set of terms and thus $P_{WTPD}{}^t$ is likely to be below $P_T{}^t$ under these conditions.

For our empirical example in the Appendix, $P_{WTPD}{}^t$ ended up being *above* $P_T{}^t$ on average rather than below: the average value for $P_{WTPD}{}^t$ for our sample was 0.99504 while the average value for $P_T{}^t$ was 0.97607. The sample averages for the two sets of covariance terms on the right hand sides of equations (150) were −0.0172 and 0.0362. Thus these covariances had the expected signs but the second set of covariances were unusually large and outweighed the negative first set of covariances, which led to $P_{WTPD}{}^t$ being above $P_T{}^t$ on average. A problem with the data for the first year is the fact that products 2 and 4 were not available for the first 8 "months" of the first year and this fact can lead to some counterintuitive results.[113] Thus when the indexes were recalculated after dropping the first year of data, $P_{WTPD}{}^t$ ended up being *below* $P_T{}^t$ on average as was expected: the average value for $P_{WTPD}{}^t$ for our truncated sample was 0.82075 while the average value for $P_T{}^t$ was 0.83100. The sample averages for the two sets of covariance terms on the right hand sides of equations (150) were −0.0195 and 0.0067.

To sum up, the weighted time product indexes can be problematic in the elementary index context with price and quantity data available when compared to a fixed base superlative index (that uses reservation prices):

- If there are no missing products and the products are strong substitutes, the indexes will tend to have a downward bias.
- If there are missing products in period 1, the indexes may have a significant upward bias.
- If there are missing products, the weighted time product dummy indexes do not depend on reservation prices.

14. Relative Price Similarity Linked Indexes

[113] The missing products in period 1 mean that s_{1n} will be equal to zero for n = 2 and 4 and thus the terms $s_{tn} - s_{1n}$ for these n will be unusually large, which in turn will usually cause the second covariance term on the right hand side of (150) to become unusually large. Remember that we are using reservation prices for the missing products; i.e., $p_{1n} > 0$ for n = 2 and 4 and these p_{1n} will tend to be greater than β_n for these n.

The GEKS multilateral method treats each set of price indexes using the prices of one period as the base period as being equally valid and hence an averaging of the parities seems to be appropriate under this hypothesis. Thus the method is "democratic" in that each bilateral index number comparison between any two periods gets the same weight in the overall method. However, it is not the case that all bilateral comparisons of price between two periods are equally accurate: if the relative prices in periods r and t are very similar, then the Laspeyres and Paasche price indexes will be very close to each other and hence it is likely that the "true" price comparison between these two periods (using the economic approach to index number theory) will be very close to the bilateral Fisher index between these two periods. In particular, if the two price vectors are exactly proportional, then we want the price index between these two periods to be equal to the factor of proportionality and the direct Fisher index between these two periods satisfies this proportionality test. On the other hand, the GEKS index comparison between the two periods would not in general satisfy the proportionality test. The above considerations suggest that a more accurate set of price indexes could be constructed if initially a bilateral comparison was made between the two countries which had the most *similar relative price structures*. At the next stage of the comparison, look for a third period which had the most similar relative price structure to the first two periods and link in this third country to the comparisons of volume between the first two countries and so on. At the end of this procedure, a pathway through the periods in the window would be constructed, that minimized the sum of the relative price dissimilarity measures. In the context of making comparisons of prices across countries, this method of linking countries with the most similar structure of relative prices has been pursued by Hill (1997) (1999a) (1999b) (2009), Diewert (2009) (2013) and Hill, Rao, Shankar and Hajargasht (2017). Hill (2001) (2004) also pursued this similarity of relative prices approach in the time series context. The conclusion is that similarity linking using Fisher ideal quantity indexes as the bilateral links is an alternative to GEKS which has some advantages over it. Both methods are consistent with the economic approach to index number theory.

A key aspect of this methodology is the choice of the measure of similarity (or dissimilarity) of the relative price structures of two countries. Various measures of the similarity or dissimilarity of relative price structures have been proposed by Allen and Diewert (1981), Kravis, Heston and Summers (1982; 104-106), Hill (1997) (2009), Sergeev (2001) (2009), Aten and Heston (2009) and Diewert (2009).

In this study, we will use the following *weighted asymptotic linear index of relative price dissimilarity*, Δ_{AL}, suggested by Diewert (2009):

$$(151)\ \Delta_{AL}(p^r,p^t,q^r,q^t) \equiv \sum_{n=1}^{N} (1/2)(s_{rn}+s_{tn})\{(p_{tn}/P_F(p^r,p^t,q^r,q^t)p_{rn}) + (P_F(p^r,p^t,q^r,q^t)p_{rn}/p_{tn}) - 2\}$$

where $P_F(p^r,p^t,q^r,q^t) \equiv [p^t{\cdot}q^r p^t{\cdot}q^t/p^r{\cdot}q^r\ p^r{\cdot}q^t]^{1/2}$ is the bilateral Fisher price index linking period t to period r and p^r, q^r, s^r and p^t, q^t, s^t are the price, quantity and share vectors for periods r and s respectively. This measure turn out to be nonnegative and the bigger $\Delta_{AL}(p^r,p^t,q^r,q^t)$ is, the more dissimilar are the relative prices for periods r and t. Note that if $p^t = \lambda p^r$ for some positive scalar so that prices are proportional for the two periods, then $\Delta_{AL}(p^r,p^t,q^r,q^t) = 0$. Note also that all prices need to be positive in order for $\Delta_{AL}(p^r,p^t,q^r,q^t)$ to be well defined. This is not a problem for our empirical example since we used estimated reservation prices for the missing prices in our sample. In the following section, we will replace our econometrically estimated reservation prices with inflation adjusted carry forward or carry backward prices for the missing prices.

The straightforward method for constructing *Similarity Linked Fisher* price indexes proceeds as follows. Set the similarity linked price index for period 1, $P_S^1 \equiv 1$. The period 2 index is set equal to $P_F(p^1,p^2,q^1,q^2)$, the Fisher index linking the period 2 prices to the period 1 prices. Thus $P_S^2 \equiv P_F(p^1,p^2,q^1,q^2)P_S^1$. For period 3, evaluate the dissimilarity indexes $\Delta_{AL}(p^1,p^3,q^1,q^3)$ and $\Delta_{AL}(p^2,p^3,q^2,q^3)$ defined by (151). If $\Delta_{AL}(p^1,p^3,q^1,q^3)$ is the minimum of these two numbers, define $P_S^3 \equiv P_F(p^1,p^2,q^1,q^2)P_S^1$. If $\Delta_{AL}(p^2,p^3,q^2,q^3)$ is the minimum of these two numbers, define $P_S^3 \equiv P_F(p^2,p^3,q^2,q^3)P_S^2$. For period 4, evaluate the dissimilarity indexes $\Delta_{AL}(p^r,p^4,q^r,q^4)$ for r = 1,2,3. Let r^* be such that $\Delta_{AL}(p^{r*},p^4,q^{r*},q^4) = \min_r \{\Delta_{AL}(p^r,p^4,q^r,q^4; r = 1,2,3\}$.[114] Then define $P_S^4 \equiv P_F(p^{r*},p^4,q^{r*},q^4)P_S^{r*}$. Continue this process in the same manner; i.e., for period t, let r^* be such that $\Delta_{AL}(p^{r*},p^t,q^{r*},q^t) = \min_r \{\Delta_{AL}(p^r,p^t,q^r,q^t; r = 1,2,...,t-1\}$ and define $P_S^t \equiv P_F(p^{r*},p^t,q^{r*},q^t)P_S^{r*}$. This procedure allows for the construction of similarity linked indexes in real time.

We implemented the above procedure with our scanner data set and compared the resulting similarity linked index, P_S^t, to our other indexes that are based on the use of superlative indexes and the economic approach to index number theory. The comparison indexes are the fixed base Fisher and Törnqvist indexes, P_F^t and P_T^t, and the multilateral indexes, P_{GEKS}^t and P_{CCDI}^t. The sample means for these five indexes, P_S^t, P_F^t, P_T^t, P_{GEKS}^t and P_{CCDI}^t were 0.97069, 0.97434, 0.97607, 0.97414 and 0.97602. Thus on average, P_S^t was about 0.5 percentage points below P_T^t and P_{CCDI}^t and about 0.35 percentage points below P_F^t and P_{GEKS}^t. These are fairly significant differences.[115]

As we have seen in previous sections, anomalous results can sometimes be due to the fact that products 2 and 4 were missing for the first 8 months in our sample. Thus we recalculated the above 5 indexes using just the data for the last two years in our sample. The resulting sample means for the five indexes, P_S^t, P_F^t, P_T^t, P_{GEKS}^t and P_{CCDI}^t were 0.83049, 0.83094, 0.83100, 0.83100 and 0.83100. Thus on average, P_S^t was only about 0.05 percentage point below the comparison indexes, which were tightly clustered.[116]

There is a problem with the above "real time" implementation of the similarity linking methodology: for the first year of observations, the allowable set of bilateral links is highly restricted. Thus the method, which relies on linking price vectors which are as close to being proportional as possible, could be generalized to allow for a preliminary phase where the first year or so of data is linked *simultaneously* as is done when making international comparisons. Thus for our *Modified Similarity Linked Fisher* price indexes,

---

[114] If the minimum occurs at more than one r, choose r^* to be the lowest of this minimizing periods.
[115] The final values for the 5 indexes were as follows: 0.92575, 0.95071, 0.95482, 0.94591 and 0.94834.
[116] The final values for the 5 indexes were as follows: 0.83048, 0.84277, 0.84144, 0.84470 and 0.85513.

$P_{SM}{}^t$, we first constructed a *Phase 1 minimum spanning tree* that linked all of the months in year 1 of our sample plus one additional month.[117] Once these indexes were defined for our first year plus one month of data, for *Phase 2*, the remaining links for $P_{SM}{}^t$ were constructed in a real time fashion as in the original method.[118] Of course, the Phase 1 indexes cannot be implemented in real time (and this is a minor disadvantage of the Modified Method) but the Phase 2 indexes can of course be implemented in real time.

We implemented the above procedure with our scanner data set and compared the resulting modified real time similarity linked index, $P_{SM}{}^t$, to its real time counterpart, $P_S{}^t$ The sample means for these two indexes were 0.96464 and 0.97069 respectively. Thus on average, $P_{SM}{}^t$ was about 0.6 percentage points below $P_S{}^t$. However, the two indexes ended up at the same level at the end of the sample period, 0.92575. This can happen because the links used by the two indexes are the same starting from month 2 in year 2 and it turns out that about ½ of the links for the first year remained unchanged and thus later indexes linked to these initial identical links will generate identical indexes for subsequent periods.[119]

As a further modification of the Modified Similarity Linking method, we looked at the effects of changing the measure of relative price dissimilarity to the following *modified weighted asymptotic linear index of relative price dissimilarity* that compares the prices of period t relative to the prices of the base period r, $\Delta_{ALM}(p^r,p^t,q^r,q^t)$, defined as follows:

$$(152)\ \Delta_{ALM}(p^r,p^t,q^r,q^t) \equiv \sum_{n=1}^N s_{tn}\{(p_{tn}/P_F(p^r,p^t,q^r,q^t)p_{rn}) + (P_F(p^r,p^t,q^r,q^t)p_{rn}/p_{tn}) - 2\}.$$

The difference between the measures of relative price dissimilarity defined by (151) and (152) is that the original measure (151) uses the average share weights of the base and current period, $(1/2)(s_{rn}+s_{tn})$, whereas the modified measure uses only the share weights of the current period, s_{tn}. It seems appropriate to use the symmetric dissimilarity measure defined by (151) in Phase 1 of the modified similarity linking procedure (thus leading to indexes that are invariant to the choice of the base period for the Phase 1 observations) but the use of the measure defined by (152) may be appropriate in Phase 2, since the focus is on linking the prices of the current period t to past prices of period $r < t$ (which introduces an asymmetric perspective) and thus the current period share weights should be given a more prominent role than past weights which may not be representative for the

[117] If there are seasonal products in our sample, this allows the prices in month 1 of the first year to be linked directly to the prices of month 1 in the second year of our sample. Our "months" consisted of 4 consecutive weeks of data which means there were 13 "months" in each of our 3 years of data. The minimum spanning tree for the first 14 "months" of our sample consisted of the following bilateral Fisher price index links: 1-12; 1-8; 8-9; 9-11; 9-13; 9-3; 3-4; 3-2; 2-14; 14-7; 14-10; 10-6; 6-5.
[118] Thus for periods $t > 14$, let r^* be such that $\Delta_{AL}(p^{r*},p^t,q^{r*},q^t) = \min_r \{\Delta_{AL}(p^r,p^4,q^r,q^4);\ r = 1,2,...,t-1\}$ and define $P_{SM}{}^t \equiv P_F(p^{r*},p^t,q^{r*},q^t)P_{SM}{}^{r*}$.
[119] We calculated the Similarity Linked and Modified Similarity Linked Fisher price indexes after dropping the data for year 1 in our sample. The resulting means for $P_{SM}{}^t$ and $P_S{}^t$ were 0.83014 and 0.83049. Thus the modified indexes were below their real time counterpart indexes by 0.035 percentage points on average. $P_S{}^t$ and P_{SM} ended up at 0.83048 and 0.82116 respectively. Thus using the Modified Method can make a small but significant difference.

current period. We implemented this modification of the modified method but the use of the new measure of relative price dissimilarity defined by (152) did not lead to a change in the pattern of bilateral links in Phase 2 of the new modified procedure and thus the new modified indexes coincided with the old modified indexes.

It appears that the similarity linked indexes, P_S^t and P_{SM}^t, generate lower rates of inflation than their superlative index counterparts, P_F^t, P_T^t, P_{GEKS}^t and P_{CCDI}^t and that there is a tendency for P_{SM}^t to be below P_S^t.[120]

What are some of the advantages and disadvantages of using either P_S^t, P_{SM}^t, P_F^t, P_T^t, P_{GEKS}^t or P_{CCDI}^t as target indexes for an elementary index in a CPI? All of these indexes are equally consistent with the economic approach to index number theory. The problem with the fixed base Fisher and Törnqvist indexes is that they depend too heavily on the base period. Moreover, sample attrition means that the base must be changed fairly frequently leading to a potential chain drift problem. The GEKS and CCDI indexes also suffer from problems associated with the existence of seasonal products: it makes little sense to include bilateral indexes between all possible periods in a window of periods in the context of seasonal commodities. The similarity linked indexes address both the problem of sample attrition and the problem of seasonal commodities. Moreover, Walsh's multiperiod identity test is always satisfied using this methodology.

However, similarity linked price indexes suffer from at least two problems:

- A measure of price dissimilarity must be chosen and there may be many "reasonable" choices of a measure which can lead to uncertainty about the choice of the measure and hence uncertainty about the resulting indexes.
- The measures of weighted price dissimilarity suggested in Diewert (2009) require that all prices in the comparison of prices between two periods be positive.

Hopefully, over time, price statisticians will come to a consensus on what "standard" relative price dissimilarity measure should be used and the first problem will be addressed. In order to address the second problem, one could use econometric methods to estimate reservation prices for products that are not present in some periods as we have done for our empirical example. But it is a difficult econometric exercise to estimate reservation prices and so a simpler method is required in order to construct imputed prices for missing products in a scanner data set. In the following section, we suggest such a method.

[120] Recall the identity (65) in section 7 above. In that section, we argued that it is likely that the direct Törnqvist index is likely to have an upward bias relative to its chained counterpart if the products are highly substitutable. The same property is likely to hold if we use Fisher indexes in place of the Törnqvist index. Thus it is likely that the similarity linked indexes using a superlative index number formula will yield a lower measure of inflation than any fixed base superlative index and hence, the similarity linked indexes using a superlative index number formula are likely to generate lower rates of price inflation than the multilateral GEKS and CCDI indexes. The same logic explains why the modified similarity linked indexes will generate index levels that are equal to or less than the corresponding unmodified real time similarity linked indexes.

15. Inflation Adjusted Carry Forward and Backward Imputed Prices

When constructing elementary indexes, statistical agencies often encounter situations where a product in an elementary index disappears. At the time of disappearance, it is unknown whether the product is temporarily unavailable so typically, the missing price is set equal to the last available price; i.e., the missing price is replaced by a carry *forward price*. Thus carry forward prices are used in place of reservation prices, which are much more difficult to construct. An alternative to the use of a carry forward price is an *inflation adjusted carry forward price*; i.e., the last available price is escalated at using the maximum overlap index between the period when the product was last available and the current period where an appropriate index number formula is used.[121] In this section, we use inflation adjusted carry forward and carry backward prices in place of the reservation prices for our scanner data set and compare the resulting indexes with our earlier indexes that used the econometrically estimated reservation prices that were constructed by Diewert and Feenstra (2017) for our data set.

Products 2 and 4 were missing from our scanner data set for observations 1-8. Let $P_{FO}(r,t)$ be the maximum overlap[122] bilateral Fisher price index for month t relative to the base month r. Define the *inflation adjusted carry backward prices* for products 2 and 4 for the missing months, p_{rn}, as follows for r = 1,2,...,8 and n = 2,4 as follows:

(153) $p_{rn} \equiv p_{9n}/P_{FO}(r,9)$.

Product 12 was missing from our sample for months 10 and 20-22. Define the *inflation adjusted carry forward price* for product 12 for the missing month 10 as follows:

(154) $p_{10,12} \equiv p_{9,12}P_{FO}(9,10)$.

Define the *inflation adjusted carry forward prices* for product 12 for the missing months 20-22 as follows for t = 20, 21, 22:

(155) $p_{t,12} \equiv p_{19,12}P_{FO}(19,t)$.

For our scanner data set, the means for the 8 reservation prices for product 2 and 4 were 0.13500 and 0.14642 respectively while the corresponding means for the carry backward prices were 0.16339 and 0.15670. Thus on average, the carry backward prices were *higher* than the reservation prices for these products, contrary to our expectations. All prices are in dollars per ounce.

The mean of the 4 reservation prices for product 12 was 0.0862 while the corresponding mean for the carry forward prices was 0.0838. Thus for product 12, on average the carry

[121] Triplett (2004; 21-29) calls these two methods for replacing missing prices the *link to show no change method* and the *deletion method*. See Diewert, Fox and Schreyer (2017) for a more extensive discussion on the problems associated with finding replacements for missing prices.

[122] A maximum overlap index is a bilateral index that is defined using only the products that are present in both periods being compared.

forward imputed prices were *lower* than the corresponding econometrically determined reservation prices, which is more in line with our expectations.[123]

Here are the sample averages for the unweighted Jevons, Dutot, fixed base Carli and chained Carli for our data set using inflation adjusted carry forward and backward prices (with the corresponding averages using reservation prices in parentheses): $P_J^\bullet$ = 0.9420 (0.9496); $P_D^\bullet$ = 0.9361 (0.9458); $P_C^\bullet$ = 0.9565 (0.9628); $P_{CCh}^\bullet$ = 1.1544 (1.1732).[124] Thus there are some significant differences between the unweighted carry forward and reservation price indexes.

We turn now to the differences in weighted indexes. A priori, we expect smaller differences in the carry forward and reservation price weighted indexes because the share of sales of frozen juice products for the missing products is quite small on average.[125] The sample averages for the fixed base Fisher, Törnqvist, Laspeyres, Paasche, Geometric Laspeyres, Geometric Paasche, and Unit Value indexes for our data set using inflation adjusted carry forward and backward prices (with the corresponding averages using reservation prices in parentheses) are as follows: $P_F^\bullet$ = 0.9716 (0.9743); $P_T^\bullet$ = 0.9738 (0.9761); $P_L^\bullet$ = 1.0430 (1.0431); $P_P^\bullet$ = 0.9067 (0.9117); $P_{GL}^\bullet$ = 1.0296 (1.0297); $P_{GP}^\bullet$ = 0.9221 (0.9263); $P_{UV}^\bullet$ = 1.0126 (1.0126).[126] Thus there are some differences between the carry forward indexes and the corresponding reservation price indexes but they are not as large as was the case for the unweighted index comparisons.

We turn our attention to the 6 multilateral indexes that we considered above. The sample averages for the GEKS, CCDI, GK, Weighted Time Product Dummy, Similarity Linked and Modified Similarity Linked indexes using inflation adjusted carry forward and backward prices (with the corresponding averages using reservation prices in parentheses) are as follows: $P_{GEKS}^\bullet$ = 0.9714 (0.9742); $P_{CCDI}^\bullet$ = 0.9737 (0.9760); $P_{GK}^\bullet$ = 0.9976 (0.9976); $P_{WTPD}^\bullet$ = 0.9950 (0.9950); $P_S^\bullet$ = 0.9689 (0.9707); $P_{SM}^\bullet$ = 0.9689 (0.9646).[127] Thus on average, use of inflation adjusted carry forward and carry backward prices compared to the use of reservation prices led to no change in P_{GK}^t and P_{WTPD}^t, lower average prices for P_{GEKS}^t, P_{CCDI}^t and P_S^t and higher average prices for P_{SM}^t. Our tentative conclusion here is that *for products that are highly substitutable, the use of*

[123] Reservation prices should be higher than the average of the prices of products that are actually available in order to induce purchasers to buy 0 units of the unavailable product.

[124] The carry forward index values for the 4 indexes for the final month in our sample were as follows: 0.91712 (0.92689); 0.91057 (0.92334); 0.94292 (0.95090); 1.40769 (1.44006) with the corresponding reservation price indexes in parentheses.

[125] The sample sales shares of products 2, 4 and 12 are 0.0213, 0.0275 and 0.0248 respectively.

[126] Recall that the unit value price index does not depend on the reservation prices. The carry forward index values for the 7 indexes for the final month in our sample were as follows: 0.9480 (0.9507); 0.9528 (0.9548); 1.0403 (1.0403); 0.8638 (0.8689); 1.0168 (1.0168); 0.8923 (0.8966); 1.0000 (1.0000) with the corresponding reservation price indexes in parentheses.

[127] Recall that P_{GK}^t and P_{WTPD}^t do not depend on the reservation prices. The carry forward index values for the 6 multilateral indexes P_{GEKS}^t, P_{CCDI}^t, P_{GK}^t, P_{WTPD}^t, P_S^t and P_{SM}^t, for the final month in our sample were as follows: 0.9425 (0.9459); 0.9455 (0.9483); 0.9616 (0.9616); 0.9645 (0.9645); 0.9294 (0.9258); 0.9294 (0.9258) with the corresponding reservation price indexes in parentheses. Thus at the end of the sample period, P_{GK}^t and P_{WTPD}^t were about 2 percentage points above P_{GEKS}^t and P_{CCDI}^t and these indexes ended up about 2 percentage points above P_S^t and P_{SM}^t. Thus the choice of index number formula matters!

carry inflation adjusted forward and backward prices for missing products will probably generate weighted indexes that are comparable to their counterparts that use econometrically estimated reservation prices.[128] This conclusion is only tentative and further research on the use of reservation prices is required. It is not easy to estimate reservation prices and the estimated reservation prices are subject to a considerable amount of sampling error and so the ability to substitute inflation adjusted carry forward and backward prices for reservation prices without substantially impacting the indexes is an important result.

When only price information is available, it is not clear that the above methodology can be adapted to this situation where quantity and value data are not available. Until more research is done, the use of the Time Product Dummy model is recommended for this situation.

16. Conclusion

Some of the more important results in each section of the paper will be summarized here.

- If there are diverging trends in product prices, the Dutot index is likely to have an upward bias relative to the Jevons index; see section 2.
- The Carli index is likely to have an upward bias relative to the Jevons index. The same result holds for the weighted Carli (or Young) index relative to the corresponding weighted Jevons index; see section 3.
- The useful relationship (41) implies that the Fisher index P_F^t will be slightly less than the corresponding fixed base Törnqvist index P_T^t, provided that the products in scope for the index are highly substitutable and there are diverging trends in prices; see section 4. Under these circumstances, the following inequalities between the Paasche, Geometric Paasche, Törnqvist, Geometric Laspeyres and Laspeyres indexes are likely to hold: $P_P^t < P_{GP}^t < P_T^t < P_{GL}^t < P_L^t$.
- The covariance identity (48) provides an exact relationship between the Jevons and Törnqvist indexes. Some conditions for equality and for divergence between these two indexes are provided at the end of section 5.
- In section 6, a geometric index that uses annual expenditure sales of a previous year as weights, $P_{J\alpha}^t$, is defined and compared to the Törnqvist index, P_T^t. Equation (61) provides an exact covariance decomposition of the difference between these two indexes. If the products are highly substitutable and there are diverging trends in prices, then it is likely that $P_T^t < P_{J\alpha}^t$.
- Section 7 derives an exact relationship (65) between the fixed base Törnqvist index, P_T^t, and its chained counterpart, P_{TCh}^t. This identity is used to show that it is likely that the chained index will "drift" below its fixed base counterpart if the products in scope are highly substitutable and prices are frequently heavily discounted. However, a numerical example shows that if quantities are slow to adjust to the lower prices, then upward chain drift can occur.

[128] As we have seen, this conclusion did not hold for unweighted indexes.

- Section 8 introduces two multilateral indexes, P_{GEKS}^t and P_{CCDI}^t. The exact identity (78) for the difference between P_{CCDI}^t and P_T^t is derived. This identity and the fact that P_F^t usually closely approximates P_T^t lead to the conclusion (79) that typically, P_F^t, P_T^t, P_{GEKS}^t and P_{CCDI}^t will approximate each other fairly closely.
- Section 9 introduces the Unit Value price index, P_{UV}^t, and shows that if there are diverging trends in prices, it is likely that $P_{UV}^t < P_F^t$. However, this conclusion does not necessarily hold if there are missing products in period 1. Section 10 derives similar results for the Quality Adjusted Unit Value index, $P_{UV\alpha}^t$.
- Section 11 looks at the relationship of the Lowe index, P_{Lo}^t, with other indexes. The Lowe index uses the quantities in a base year as weights in a fixed basket type index for months that follow the base year. In using annual weights of a previous year, this index is similar in spirit to the geometric index $P_{J\alpha}^t$ that was analyzed in section 6. The covariance type identities (128) and (131) are used to suggest that it is likely that the Lowe index lies between the fixed base Paasche and Laspeyres indexes; i.e., it is likely that $P_P^t < P_{Lo}^t < P_L^t$. The identity (134) is used to suggest that the Lowe index is likely to have an upward bias relative to the fixed base Fisher index; i.e., it is likely that $P_F^t < P_{Lo}^t$. However, if there are missing products in the base year, then these inequalities do not necessarily hold.
- Section 12 looks at an additional multilateral index, the Geary Khamis index, P_{GK}^t and shows that P_{GK}^t can be interpreted as a quality adjusted unit value index and hence using the analysis in section 10, it is likely that the Geary Khamis index has a downward bias relative to the Fisher index; i.e., it is likely that $P_{GK}^t < P_F^t$. However, if there are missing products in the first month of the sample, the above inequality will not necessarily hold.
- Another multilateral index is introduced in section 13: the Weighted Time Product Dummy index, P_{WTPD}^t. The exact identity (150) is used to show that it is likely that P_{WTPD}^t is less than the corresponding fixed base Fisher index, P_F^t, provided that the products are highly substitutable and there are no missing products in period 1.
- It turns out that the following price indexes are not affected by reservation prices: P_{UV}^t, $P_{UV\alpha}^t$, P_{GK}^t and P_{WTPD}^t. Thus these indexes are not consistent with economic approach to dealing with the problems associated with new and disappearing products and services.
- The final multilateral index is introduced in section 14: Similarity Linked price indexes. A real time version of the method is introduced, P_S^t, along with a modified version, P_{SM}^t, that requires a year of “training” data before the introduction of the real time indexes.
- For our empirical example, P_S^t and P_{SM}^t ended up about 2 percentage points below P_{GEKS}^t and P_{CCDI}^t which in turn finished about 1 percentage point below P_F^t and P_T^t and finally P_{GK}^t and P_{WTPD}^t finished about 1 percentage point above P_F^t and P_T^t; see Chart 8 in the Appendix. All 8 multilateral indexes captured the trend in product prices quite well. More research is required in order to determine whether these differences are significant.
- It is difficult to calculate reservation prices using econometric techniques. Thus section 15 looked at methods for replacing reservation prices by inflation adjusted carry forward and backward prices which are much easier to calculate.

- For our empirical example, the replacement of the reservation prices by carried prices did not make much difference to the multilateral indexes. If the products in scope are highly substitutable for each other, then we expect that this invariance result will hold (approximately). However, if products with new characteristics are introduced, then we expect that the replacement of econometrically estimated reservation prices by carried prices would probably lead to an index that has an upward bias.

Conceptually, the Modified Similarity linked indexes seem to be the most attractive solution for solving the chain drift problem since they can deal with seasonal commodities and as well, Walsh's multiperiod identity test will always be satisfied.

The data used for the empirically constructed indexes are listed in the Appendix so that the listed indexes can be replicated and so that alternative solutions to the chain drift problem can be tested out by other statisticians and economists.

Appendix: Data Listing and Index Number Tables and Charts

A.1 Data Listing

Here is a listing of the "monthly" quantities sold of 19 varieties of frozen juice (mostly orange juice) from Dominick's Store 5 in the Greater Chicago area, where a "month" consists of sales for 4 consecutive weeks. The weekly unit value price and quantity sold data were converted into "monthly" unit value prices and quantities.[129] Finally, the original data came in units where the package size was not standardized. We rescaled the price and quantity data into prices per ounce. Thus the quantity data are equal to the "monthly" ounces sold for each product.

Table A1: "Monthly" Unit Value Prices for 19 Frozen Juice Products

t	p_1^t	p_2^t	p_3^t	p_4^t	p_5^t	p_6^t	p_7^t	p_8^t	p_9^t
1	0.122500	*0.145108*	0.147652	*0.148593*	0.146818	0.146875	0.147623	0.080199	0.062944
2	0.118682	*0.127820*	0.116391	*0.128153*	0.117901	0.146875	0.128833	0.090833	0.069167
3	0.120521	*0.128608*	0.129345	*0.148180*	0.131117	0.143750	0.136775	0.090833	0.048803
4	0.126667	*0.128968*	0.114604	*0.115604*	0.116703	0.143750	0.114942	0.088523	0.055842
5	0.126667	*0.130737*	0.140833	*0.141108*	0.140833	0.143304	0.140833	0.090833	0.051730
6	0.120473	*0.113822*	0.157119	*0.151296*	0.156845	0.161844	0.156342	0.090833	0.049167
7	0.164607	*0.144385*	0.154551	*0.158485*	0.156607	0.171875	0.152769	0.084503	0.069167
8	0.142004	*0.160519*	0.174167	*0.179951*	0.174167	0.171341	0.163333	0.089813	0.069167
9	0.135828	0.165833	0.154795	0.159043	0.151628	0.171483	0.160960	0.089970	0.067406
10	0.129208	0.130126	0.153415	0.158167	0.152108	0.171875	0.158225	0.078906	0.067897
11	0.165833	0.165833	0.139690	0.136830	0.134743	0.171875	0.136685	0.079573	0.058841
12	0.165833	0.165833	0.174167	0.174167	0.174167	0.171875	0.174167	0.081902	0.079241
13	0.113739	0.116474	0.155685	0.149942	0.145633	0.171875	0.146875	0.074167	0.048880

[129] In practice, statistical agencies will not be able to produce indexes for 13 months in a year. There are two possible solutions to the problem of aggregating weekly data into monthly data: (i) aggregate the data for the first 3 weeks in a month or (ii) split the weekly data that spans two consecutive months into imputed data for each month.

14	**0.120882**	**0.125608**	**0.141602**	**0.147428**	**0.142664**	**0.163750**	**0.144911**	**0.090833**	**0.080000**
15	**0.165833**	**0.165833**	**0.147067**	**0.143214**	**0.144306**	**0.155625**	**0.147546**	**0.088410**	**0.080000**
16	**0.122603**	**0.118536**	**0.135878**	**0.137359**	**0.137480**	**0.155625**	**0.138146**	**0.084489**	**0.080000**
17	**0.104991**	**0.104659**	**0.112497**	**0.113487**	**0.110532**	**0.141250**	**0.113552**	**0.082500**	**0.067104**
18	**0.088056**	**0.091133**	**0.118440**	**0.120331**	**0.117468**	**0.141250**	**0.124687**	**0.085000**	**0.065664**
19	**0.096637**	**0.097358**	**0.141667**	**0.141667**	**0.141667**	**0.141250**	**0.141667**	**0.082500**	**0.080000**
20	**0.085845**	**0.090193**	**0.120354**	**0.122168**	**0.113110**	**0.136250**	**0.124418**	**0.085874**	**0.051003**
21	**0.094009**	**0.100208**	**0.121135**	**0.122500**	**0.121497**	**0.125652**	**0.121955**	**0.090833**	**0.085282**
22	**0.084371**	**0.087263**	**0.120310**	**0.123833**	**0.118067**	**0.125492**	**0.124167**	**0.085898**	**0.063411**
23	**0.123333**	**0.123333**	**0.116412**	**0.118860**	**0.113085**	**0.126250**	**0.118237**	**0.085891**	**0.049167**
24	**0.078747**	**0.081153**	**0.125833**	**0.125833**	**0.125833**	**0.126250**	**0.125833**	**0.090833**	**0.049167**
25	**0.088284**	**0.092363**	**0.098703**	**0.098279**	**0.088839**	**0.126250**	**0.100640**	**0.090833**	**0.049167**
26	**0.123333**	**0.123333**	**0.092725**	**0.096323**	**0.095115**	**0.126250**	**0.095030**	**0.090833**	**0.049167**
27	**0.101331**	**0.102442**	**0.125833**	**0.125833**	**0.125833**	**0.126250**	**0.125833**	**0.090833**	**0.049167**
28	**0.101450**	**0.108416**	**0.092500**	**0.097740**	**0.091025**	**0.126250**	**0.096140**	**0.054115**	**0.049167**
29	**0.123333**	**0.123333**	**0.118986**	**0.119509**	**0.115603**	**0.126250**	**0.118343**	**0.096922**	**0.049167**
30	**0.094038**	**0.095444**	**0.109096**	**0.113827**	**0.106760**	**0.126250**	**0.113163**	**0.089697**	**0.049167**
31	**0.130179**	**0.130000**	**0.110257**	**0.115028**	**0.112113**	**0.134106**	**0.110579**	**0.093702**	**0.049167**
32	**0.103027**	**0.103299**	**0.149167**	**0.149167**	**0.149167**	**0.149375**	**0.149167**	**0.098333**	**0.049167**
33	**0.148333**	**0.148333**	**0.089746**	**0.097110**	**0.091357**	**0.149375**	**0.094347**	**0.098333**	**0.049167**
34	**0.115247**	**0.114789**	**0.123151**	**0.123892**	**0.127177**	**0.149375**	**0.125362**	**0.094394**	**0.049167**
35	**0.118090**	**0.120981**	**0.121191**	**0.129477**	**0.128180**	**0.149375**	**0.132934**	**0.096927**	**0.049167**
36	**0.132585**	**0.131547**	**0.129430**	**0.128314**	**0.121833**	**0.134375**	**0.128874**	**0.070481**	**0.049167**
37	**0.114056**	**0.115491**	**0.138214**	**0.140090**	**0.139116**	**0.146822**	**0.142770**	**0.077785**	**0.053864**
38	**0.142500**	**0.142500**	**0.134677**	**0.133351**	**0.133216**	**0.148125**	**0.132873**	**0.108333**	**0.054167**
39	**0.121692**	**0.123274**	**0.095236**	**0.102652**	**0.093365**	**0.148125**	**0.101343**	**0.090180**	**0.054167**

t	p_{10}^t	p_{11}^t	p_{12}^t	p_{13}^t	p_{14}^t	p_{15}^t	p_{16}^t	p_{17}^t	p_{18}^t	p_{19}^t
1	**0.062944**	**0.075795**	**0.080625**	**0.087684**	**0.109375**	**0.113333**	**0.149167**	**0.122097**	**0.149167**	**0.124492**
2	**0.069167**	**0.082500**	**0.080625**	**0.112500**	**0.109375**	**0.113333**	**0.119996**	**0.109861**	**0.130311**	**0.117645**
3	**0.043997**	**0.082500**	**0.078546**	**0.106468**	**0.100703**	**0.110264**	**0.134380**	**0.109551**	**0.131890**	**0.114933**
4	**0.055705**	**0.082500**	**0.080625**	**0.099167**	**0.099375**	**0.111667**	**0.109005**	**0.106843**	**0.108611**	**0.118333**
5	**0.051687**	**0.071670**	**0.080625**	**0.094517**	**0.099375**	**0.111667**	**0.105168**	**0.106839**	**0.105055**	**0.076942**
6	**0.049167**	**0.078215**	**0.080625**	**0.115352**	**0.114909**	**0.130149**	**0.099128**	**0.134309**	**0.118647**	**0.088949**
7	**0.069167**	**0.069945**	**0.080625**	**0.124167**	**0.118125**	**0.131667**	**0.102524**	**0.128471**	**0.102073**	**0.160833**
8	**0.069167**	**0.082500**	**0.080625**	**0.107381**	**0.121513**	**0.138184**	**0.164245**	**0.141978**	**0.164162**	**0.136105**
9	**0.067401**	**0.082500**	**0.074375**	**0.112463**	**0.128125**	**0.141667**	**0.163333**	**0.153258**	**0.163333**	**0.118979**
10	**0.067688**	**0.082500**	***0.100545***	**0.132500**	**0.128125**	**0.141667**	**0.133711**	**0.152461**	**0.133806**	**0.118439**
11	**0.060008**	**0.082500**	**0.080625**	**0.120362**	**0.134151**	**0.144890**	**0.163333**	**0.151033**	**0.163333**	**0.120424**
12	**0.079325**	**0.071867**	**0.080625**	**0.093144**	**0.136875**	**0.148333**	**0.144032**	**0.148107**	**0.146491**	**0.160833**
13	**0.064028**	**0.069934**	**0.067280**	**0.118009**	**0.136875**	**0.148333**	**0.163333**	**0.143125**	**0.163333**	**0.131144**
14	**0.080000**	**0.078491**	**0.075211**	**0.131851**	**0.130342**	**0.143013**	**0.123414**	**0.152937**	**0.130223**	**0.122899**
15	**0.080000**	**0.082500**	**0.080625**	**0.093389**	**0.128125**	**0.141667**	**0.117955**	**0.147024**	**0.119786**	**0.128929**
16	**0.080000**	**0.086689**	**0.080625**	**0.100592**	**0.128125**	**0.141667**	**0.114940**	**0.143125**	**0.126599**	**0.124620**
17	**0.065670**	**0.088333**	**0.072941**	**0.115559**	**0.110426**	**0.139379**	**0.107709**	**0.143125**	**0.109987**	**0.145556**
18	**0.064111**	**0.091286**	**0.069866**	**0.088224**	**0.105625**	**0.105529**	**0.089141**	**0.130110**	**0.095463**	**0.140000**
19	**0.080000**	**0.094167**	**0.088125**	**0.080392**	**0.105625**	**0.131667**	**0.086086**	**0.118125**	**0.091020**	**0.109424**
20	**0.048613**	**0.094167**	***0.096177***	**0.080643**	**0.105625**	**0.131667**	**0.125000**	**0.114706**	**0.125000**	**0.110921**
21	**0.085114**	**0.080262**	***0.064774***	**0.080245**	**0.099375**	**0.125000**	**0.104513**	**0.114795**	**0.104228**	**0.134014**
22	**0.062852**	**0.086115**	***0.083132***	**0.087551**	**0.101493**	**0.127366**	**0.086484**	**0.118125**	**0.088325**	**0.126667**
23	**0.049167**	**0.095833**	**0.090625**	**0.089110**	**0.099375**	**0.125000**	**0.086263**	**0.118125**	**0.095750**	**0.100780**
24	**0.049167**	**0.095833**	**0.090625**	**0.090167**	**0.099375**	**0.125000**	**0.111859**	**0.114330**	**0.112296**	**0.118333**
25	**0.049167**	**0.095833**	**0.090625**	**0.072861**	**0.099375**	**0.125000**	**0.125000**	**0.113823**	**0.125000**	**0.084817**
26	**0.049167**	**0.095833**	**0.090625**	**0.086226**	**0.099375**	**0.125000**	**0.086088**	**0.114190**	**0.091864**	**0.118333**
27	**0.049167**	**0.077500**	**0.076875**	**0.081764**	**0.099375**	**0.125000**	**0.113412**	**0.114231**	**0.113241**	**0.110346**
28	**0.049167**	**0.077500**	**0.076875**	**0.104167**	**0.099375**	**0.125000**	**0.085803**	**0.118125**	**0.086154**	**0.084604**
29	**0.049167**	**0.077500**	**0.076875**	**0.086713**	**0.099375**	**0.125000**	**0.087410**	**0.118125**	**0.086196**	**0.085034**
30	**0.049167**	**0.077500**	**0.076875**	**0.104167**	**0.099375**	**0.125000**	**0.084953**	**0.114826**	**0.085156**	**0.083921**
31	**0.049167**	**0.077500**	**0.076875**	**0.095613**	**0.099375**	**0.125000**	**0.087372**	**0.125809**	**0.087775**	**0.088304**
32	**0.049167**	**0.077500**	**0.076875**	**0.112500**	**0.099375**	**0.067046**	**0.091827**	**0.143125**	**0.088937**	**0.103519**
33	**0.049167**	**0.077500**	**0.076875**	**0.104721**	**0.099375**	**0.125000**	**0.131399**	**0.143125**	**0.130253**	**0.127588**
34	**0.049167**	**0.077500**	**0.076875**	**0.088935**	**0.099375**	**0.125000**	**0.123037**	**0.143125**	**0.123573**	**0.132500**
35	**0.049167**	**0.077500**	**0.076875**	**0.112500**	**0.099375**	**0.125000**	**0.125832**	**0.137837**	**0.125681**	**0.112286**

36	**0.049167**	**0.077500**	**0.076875**	**0.089456**	**0.099375**	**0.125000**	**0.139240**	**0.141242**	**0.144390**	**0.127323**
37	**0.053865**	**0.084549**	**0.083343**	**0.107198**	**0.119368**	**0.151719**	**0.146126**	**0.154886**	**0.146332**	**0.120616**
38	**0.054167**	**0.085000**	**0.084375**	**0.127500**	**0.123125**	**0.156667**	**0.129577**	**0.138823**	**0.130850**	**0.114177**
39	**0.054167**	**0.085000**	**0.084375**	**0.102403**	**0.123125**	**0.156667**	**0.115965**	**0.149219**	**0.114947**	**0.136667**

The actual prices p_2^t and p_4^t are not available for t =1,2,...,8 since products 2 and 4 were not sold during these months. However, in the above Table, we filled in these missing prices with the imputed reservation prices that were estimated by Diewert and Feenstra (2017). Similarly, p_{12}^t was missing for months t = 12, 20, 21 and 22 and again, we replaced these missing prices with the corresponding estimated imputed reservation prices in Table A1. The imputed prices appear in italics in the above Table.

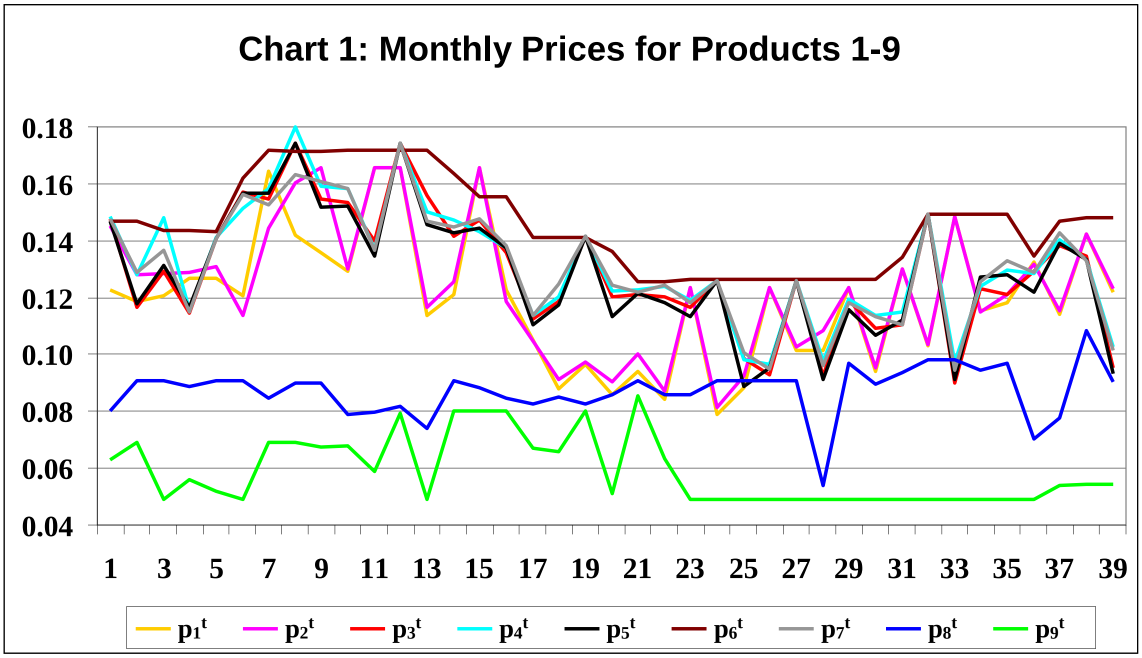
Table A2: "Monthly" Quantities Sold for 19 Frozen Juice Products

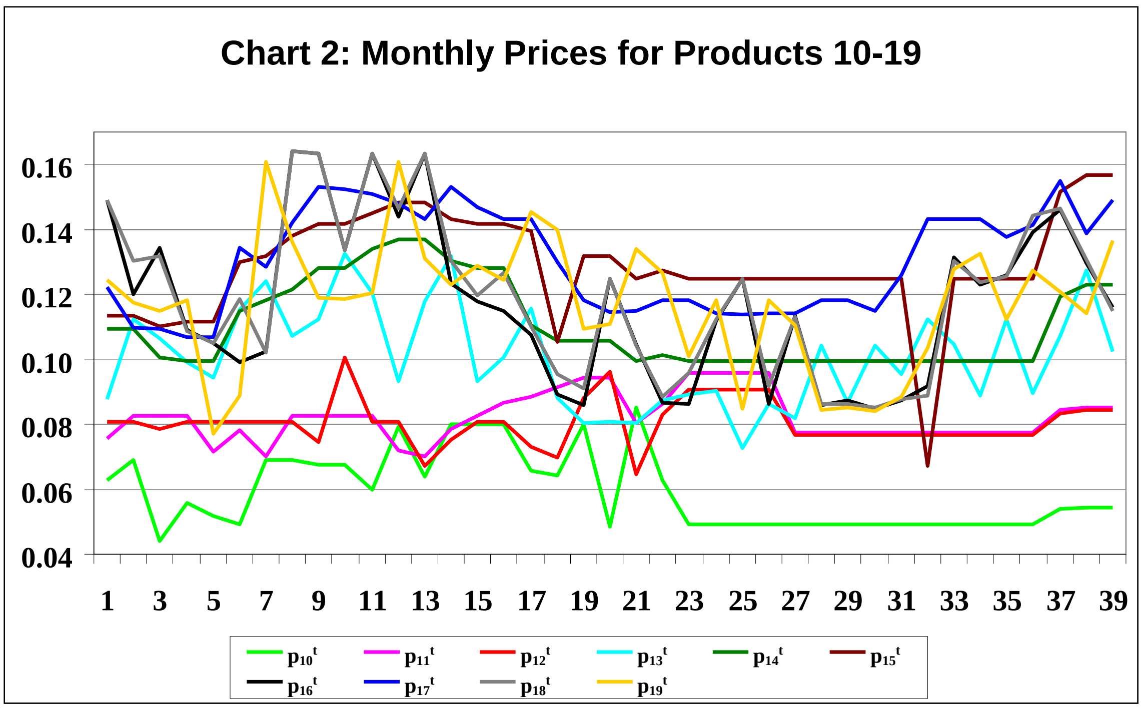
t	q_1^t	q_2^t	q_3^t	q_4^t	q_5^t	q_6^t	q_7^t	q_8^t	q_9^t
1	1704	0.000	792	0.000	4428	1360	1296	1956	1080
2	3960	0.000	3588	0.000	19344	3568	3600	2532	2052
3	5436	0.000	1680	0.000	8100	3296	2760	3000	1896
4	1584	0.000	5532	0.000	21744	3360	5160	3420	2328
5	1044	0.000	1284	0.000	5880	3360	1896	3072	1908
6	8148	0.000	1260	0.000	7860	2608	2184	3000	2040
7	636	0.000	3120	0.000	9516	2848	2784	3444	1620
8	1692	0.000	1200	0.000	4116	1872	1380	2088	1848
9	5304	1476.000	2292	1295.999	7596	2448	1740	2016	3180
10	6288	2867.993	2448	1500.000	6528	2064	2208	3840	4680
11	408	228.000	2448	2147.994	9852	2096	2700	5124	12168
12	624	384.000	948	1020.000	2916	1872	1068	2508	4032
13	6732	2964.005	1488	2064.003	8376	2224	2400	4080	8928
14	6180	3192.007	2472	2244.006	7920	1920	2256	1728	1836
15	1044	672.000	1572	1932.002	2880	1744	1728	1692	1116
16	3900	1332.002	1560	2339.997	4464	2416	2028	2112	1260
17	5328	1847.999	3528	3972.008	13524	2336	3252	2628	1524
18	7056	2100.000	2436	2748.007	6828	2544	1980	3000	1596
19	5712	3167.988	1464	1872.000	2100	2080	1572	3384	1020
20	9960	3311.996	2376	2172.003	8028	2112	1788	2460	3708
21	7368	2496.000	1992	1872.000	3708	1840	1980	1692	2232
22	9168	4835.983	2064	1980.000	10476	1504	2880	2472	7020
23	7068	660.000	1728	1955.999	6972	1888	2172	2448	12120
24	11856	5604.017	972	1464.000	2136	1296	1536	3780	7584
25	7116	2831.994	2760	2207.996	12468	1776	2580	2880	11220
26	660	504.000	3552	3755.995	17808	1296	5580	4956	7428
27	4824	3276.011	1356	1452.000	2388	1824	1524	1548	10188
28	3684	971.998	4680	2832.003	11712	1712	4308	4284	1140
29	684	1152.000	1884	2015.996	9252	1680	3144	1020	1392
30	5112	3467.996	2256	2291.994	9060	1936	2172	1452	2532
31	672	840.000	4788	2951.990	9396	1856	4644	1764	1260
32	7344	5843.997	1320	1128.000	2664	1744	1560	1548	1416
33	480	504.000	6624	5639.996	13368	1824	6888	1800	1440
34	4104	3036.001	2124	3180.009	5088	1568	2820	1668	1884
35	2688	1583.997	2220	2760.008	5244	1344	2532	1920	4956
36	936	612.001	1824	2567.994	6684	1552	2772	4740	7644
37	4140	2268.001	1932	1559.997	4740	1520	2076	1752	6336
38	912	264.000	1860	2844.002	4260	1808	2064	1452	2952
39	1068	960.001	4356	2903.996	11052	1776	4356	2220	2772

t	q_{10}^t	q_{11}^t	q_{12}^t	q_{13}^t	q_{14}^t	q_{15}^t	q_{16}^t	q_{17}^t	q_{18}^t	q_{19}^t
1	540	2088	1744.000	30972	3728	792	1512	1712	600	2460
2	1308	4212	3824.000	11796	6480	2712	12720	3312	2376	1788
3	1416	3900	4848.010	18708	10064	2652	4116	3184	1476	3756
4	1716	3156	5152.000	19656	10352	2472	15420	3120	3888	900
5	1452	6168	3360.000	42624	7360	1590	9228	2800	5652	13560
6	1068	5088	3296.000	10380	7712	1884	12012	1808	3348	7824
7	1116	6372	3712.000	11772	7920	1680	29592	3296	11712	708
8	1296	3684	3216.000	21024	5856	1206	11184	1744	4344	6036
9	2220	4512	3024.000	24420	5856	1398	2040	1648	1176	7896
10	4152	4572	0.000	8328	6384	1740	9168	1296	2832	9120
11	9732	3432	3360.000	18372	5808	1638	2412	1568	972	7176
12	3024	6132	1792.000	48648	4672	1770	7512	2208	2052	3564
13	2160	6828	6271.998	15960	4736	1662	1740	2896	1176	3216
14	1356	5088	2991.997	9432	5872	1902	4968	1488	2064	6420
15	1188	4656	2976.000	33936	3872	1452	9060	1744	2712	3876
16	816	3108	4784.000	23772	6272	1578	8496	2832	1488	4128
17	696	3252	4879.997	10656	7648	1836	9000	2704	2292	648
18	720	2940	4848.021	26604	6448	4086	14592	1552	3108	732
19	624	4320	2480.000	27192	4944	1140	19056	1808	5088	5676
20	3288	2784	0.000	23796	5120	1284	2196	2896	1260	3876
21	1848	12324	0.000	25824	5248	1140	8640	1952	2940	588
22	4824	6468	0.000	18168	3872	930	15360	1520	4728	276
23	10092	3708	1744.000	14592	4336	870	14232	1504	2040	1128
24	6372	3264	2016.000	16548	4608	858	6696	1792	2496	792
25	7284	3480	2032.000	38880	4064	750	1836	1232	636	7608
26	6588	3768	2208.000	14724	3760	768	9096	1296	4248	480
27	2832	4692	2592.000	31512	5344	930	5796	2080	5244	1416
28	900	3180	2624.000	8172	5776	810	13896	1328	7536	6744
29	1128	3948	2608.000	19440	5792	954	12360	1552	5796	7296
30	1284	5232	2960.000	6552	6320	924	13932	2304	8064	14520
31	864	5928	3280.000	16896	5888	852	14340	2064	8412	3768
32	948	5784	2496.000	5880	5088	15132	14496	1600	10440	4044
33	708	5232	2704.000	15180	4800	618	4812	976	3204	1812
34	1152	4692	2736.000	25344	5648	600	6552	1360	3876	1344
35	4248	4668	2800.000	8580	5488	498	28104	1872	11292	4152
36	6492	4872	2256.000	30276	5504	510	4080	1328	3768	1860
37	5976	3396	1743.995	8208	2832	384	1092	528	1284	2028
38	1812	3660	2416.000	4392	4144	534	4752	1504	2436	4980
39	2844	3852	1888.000	16704	3488	708	6180	1600	4236	804

It can be seen that there were no sales of Products 2 and 4 for months 1-8 and there were no sales of Product 12 in month 10 and in months 20-22.

Charts of the above prices and quantities follow.





It can be seen that there is a considerable amount of variability in these per ounce monthly unit value prices for frozen juice products. There are also differences in the average level of the prices of these 19 products. These differences can be interpreted as quality differences.

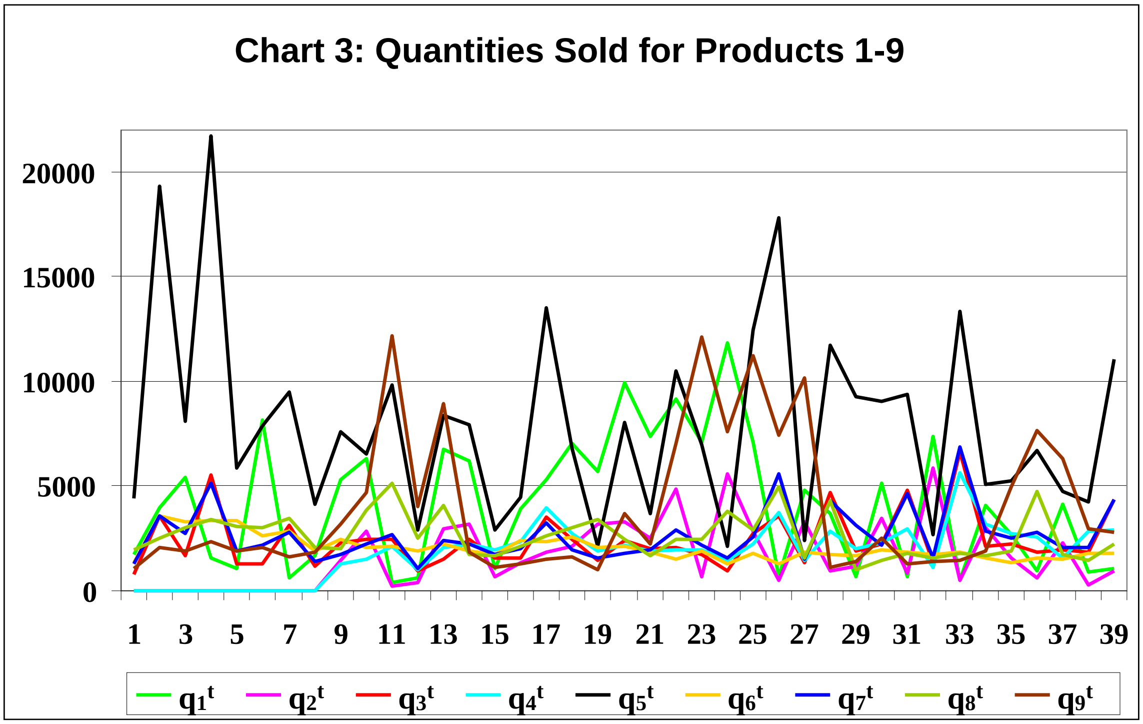


Chart 3: Quantities Sold for Products 1-9
20000
15000
10000
5000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
q1t q2t q3t q4t q5t q6t q7t q8t q9t

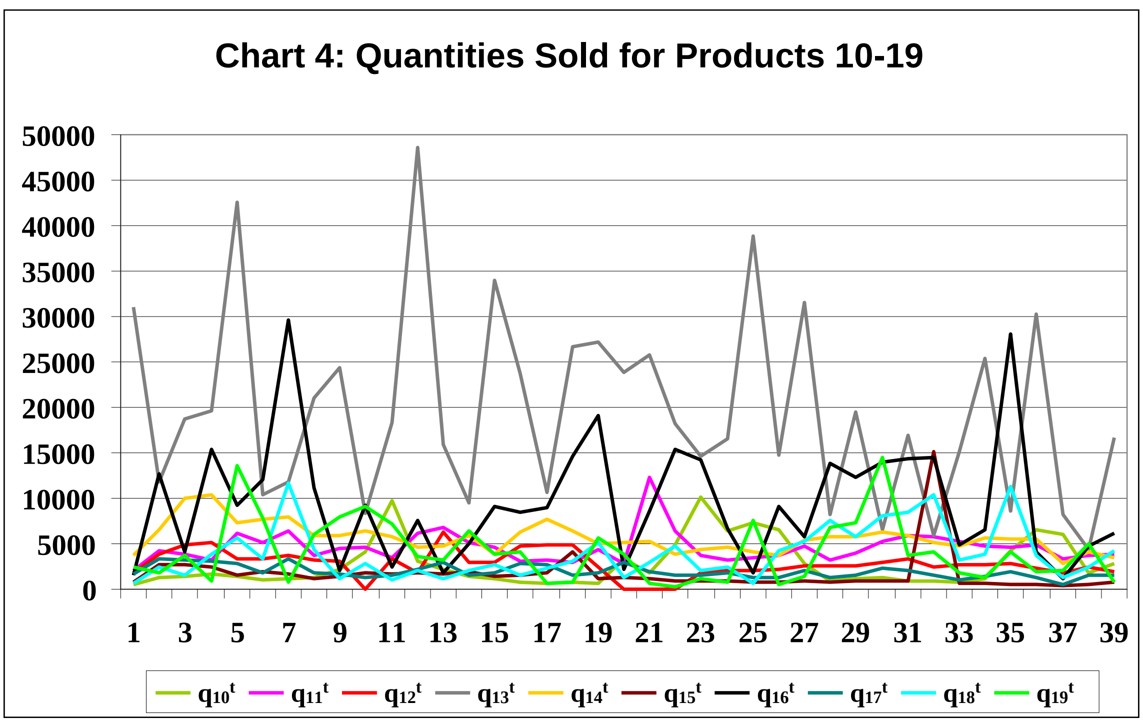


Chart 4: Quantities Sold for Products 10-19
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
q10t q11t q12t q13t q14t q15t q16t q17t q18t q19t

It can be seen that the sales volatility of the products is even bigger than the volatility in prices.

A.2 Unweighted Price Indexes

In this section, we list the unweighted indexes[130] that were defined in sections 2 and 3 in the main text. We used the data that is listed in section A.1 above in order to construct the indexes. We list the Jevons, Dutot, Carli, Chained Carli, CES with r = −1 and r = −9 which we denote by P_J^t, P_D^t, P_C^t, P_{CCh}^t, $P_{CES,-1}^t$ and $P_{CES,-9}^t$ respectively for month t.

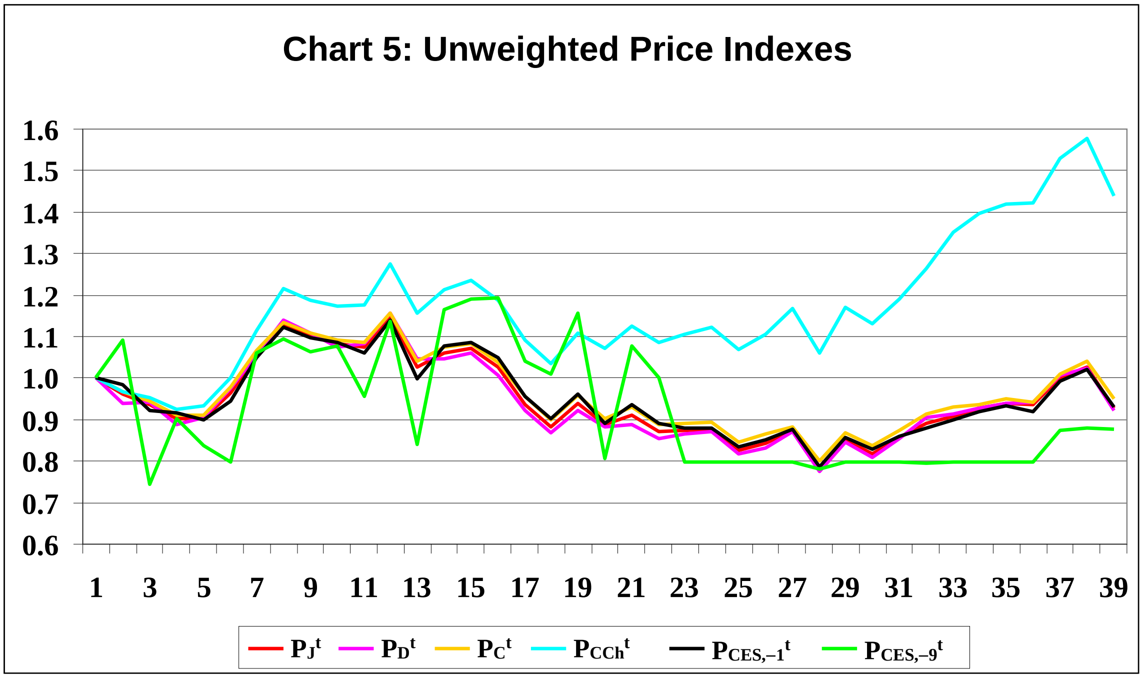
Table A.3 Jevons, Dutot, Fixed Base and Chained Carli and CES Price Indexes

t	P_J^t	P_D^t	P_C^t	P_{CCh}^t	$P_{CES,-1}$	$P_{CES,-9}$
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.96040	0.94016	0.96846	0.96846	0.98439	1.09067
3	0.93661	0.94070	0.94340	0.95336	0.92229	0.74456
4	0.90240	0.88954	0.91068	0.92488	0.91540	0.90219
5	0.90207	0.90438	0.91172	0.93347	0.89823	0.83603
6	0.96315	0.97490	0.97869	1.00142	0.94497	0.79833
7	1.05097	1.05468	1.06692	1.11301	1.04802	1.06093
8	1.13202	1.13825	1.13337	1.21622	1.12388	1.09382
9	1.10373	1.10769	1.10739	1.18820	1.09706	1.06198
10	1.08176	1.07574	1.09119	1.17299	1.08685	1.07704
11	1.07545	1.08438	1.08516	1.17758	1.06038	0.95660
12	1.14864	1.15654	1.15517	1.27479	1.13881	1.13589
13	1.02772	1.04754	1.03943	1.15786	0.99848	0.84113
14	1.06109	1.04636	1.07433	1.21248	1.07853	1.16587
15	1.07130	1.06066	1.08164	1.23459	1.08565	1.19072
16	1.02572	1.00635	1.03655	1.18788	1.04979	1.19448
17	0.93668	0.92185	0.95548	1.09129	0.95529	1.04194
18	0.88243	0.86882	0.89940	1.03405	0.90087	1.01058
19	0.93855	0.92175	0.96016	1.10908	0.96244	1.15748
20	0.88855	0.88248	0.90225	1.07164	0.89126	0.80633
21	0.91044	0.88862	0.92930	1.12593	0.93740	1.07775
22	0.87080	0.85512	0.88891	1.08595	0.89107	1.00076
23	0.87476	0.86577	0.89065	1.10508	0.88042	0.79864
24	0.87714	0.87111	0.89384	1.12219	0.87980	0.79810
25	0.82562	0.81640	0.84467	1.06793	0.83434	0.79708
26	0.84210	0.83168	0.86532	1.10572	0.85123	0.79827
27	0.87538	0.87012	0.88197	1.16687	0.87714	0.79760
28	0.78149	0.77534	0.80014	1.05919	0.78770	0.78068
29	0.85227	0.84699	0.86721	1.17131	0.85568	0.79718
30	0.81870	0.80899	0.83656	1.13006	0.82799	0.79688
31	0.85842	0.85377	0.87514	1.19113	0.86118	0.79741
32	0.89203	0.90407	0.91440	1.26420	0.87884	0.79524
33	0.90818	0.91368	0.93127	1.35047	0.89955	0.79775
34	0.92659	0.92742	0.93489	1.39685	0.91949	0.79804
35	0.93981	0.93944	0.94941	1.42023	0.93256	0.79825
36	0.93542	0.94295	0.94087	1.42210	0.92057	0.79654
37	1.00182	1.00595	1.01060	1.52914	0.99139	0.87270
38	1.02591	1.02295	1.04068	1.57788	1.02072	0.87939

[130] It would be more accurate to call these indexes equally weighted indexes.

39 0.92689 0.92334 0.95090 1.44006 0.93017 0.87789

The above prices are plotted on Chart 5.



The Chained Carli index, P_{CCh}^t, is well above the other indexes as is expected. The fixed base Carli index P_C^t is above the corresponding Jevons index P_J^t which in turn is slightly above the corresponding Dutot index P_D^t. The CES index with $r = -1$ (this corresponds to $\sigma = 2$) is on average between the Jevons and fixed base Carli indexes while the CES index with $r = -9$ (this corresponds to $\sigma = 10$) is well below all of the other indexes on average (and is extremely volatile).[131]

We turn now to a listing of standard bilateral indexes using the 3 years of data and the econometrically estimated reservation prices.

A.3 Weighted Price Indexes

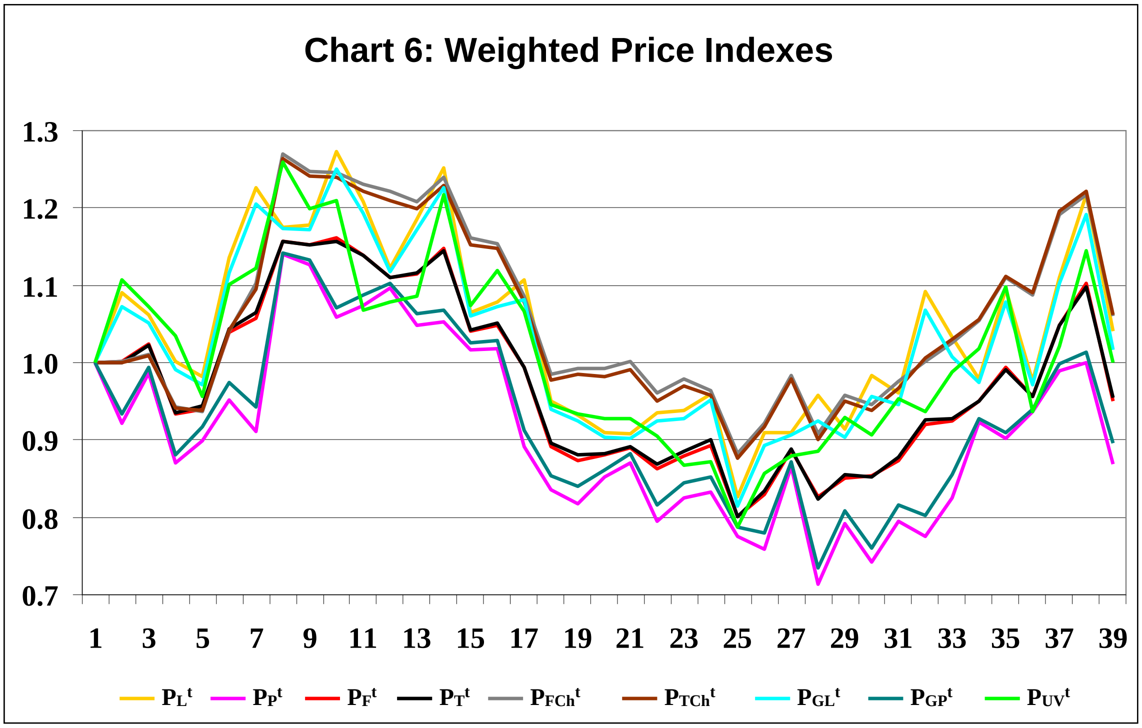
We list the fixed base and chained Laspeyres, Paasche, Fisher and Törnqvist indexes in Table A.4 below. The Geometric Laspeyres and Geometric Paasche and Unit Value indexes are also listed in this table.

Table A.4 Fixed Base and Chained Laspeyres, Paasche, Fisher and Törnqvist, Geometric Laspeyres, Geometric Paasche and Unit Value Indexes

[131] The sample means of the P_J^t, P_D^t, P_C^t, P_{CCh}^t, $P_{CES,-1}^t$ and $P_{CES,-9}^t$ are: 0.9496, 0.9458, 0.9628, 1.1732, 0.9520 and 0.9237 respectively.

t	P_L^t	P_P^t	P_F^t	P_T^t	P_{LCh}^t	P_{PCh}^t	P_{FCh}^t	P_{TCh}^t	P_{GL}^t	P_{GP}^t	P_{UV}^t
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.08991	0.92151	1.00218	1.00036	1.08991	0.92151	1.00218	1.00036	0.98194	1.07214	1.10724
3	1.06187	0.98637	1.02342	1.02220	1.12136	0.91193	1.01124	1.00905	0.97979	1.05116	1.07205
4	1.00174	0.87061	0.93388	0.93445	1.06798	0.83203	0.94265	0.94077	0.91520	0.99062	1.03463
5	0.98198	0.89913	0.93964	0.94387	1.11998	0.78417	0.93715	0.93753	0.91048	0.97176	0.95620
6	1.13639	0.95159	1.03989	1.04311	1.27664	0.84845	1.04075	1.04165	0.99679	1.11657	1.10159
7	1.22555	0.91097	1.05662	1.06555	1.42086	0.85482	1.10208	1.09531	1.07355	1.20485	1.12167
8	1.17447	1.14057	1.15740	1.15743	1.75897	0.91677	1.26987	1.26340	1.14865	1.17300	1.25911
9	1.17750	1.12636	1.15164	1.15169	1.73986	0.89414	1.24727	1.24135	1.12700	1.17162	1.19939
10	1.27247	1.05895	1.16081	1.15735	1.80210	0.86050	1.24528	1.23902	1.12514	1.25074	1.20900
11	1.20770	1.07376	1.13876	1.13875	1.86610	0.81117	1.23034	1.22114	1.12189	1.19276	1.06812
12	1.12229	1.09688	1.10951	1.10976	2.01810	0.73863	1.22091	1.20993	1.12209	1.11767	1.07795
13	1.18583	1.04861	1.11511	1.11677	2.17862	0.66995	1.20813	1.19943	1.09272	1.17231	1.08595
14	1.25239	1.05236	1.14803	1.14485	2.30844	0.66552	1.23948	1.22942	1.09463	1.22682	1.21698
15	1.06527	1.01701	1.04086	1.04292	2.32124	0.58025	1.16056	1.15215	1.03397	1.06020	1.07438
16	1.07893	1.01866	1.04836	1.05073	2.34342	0.56876	1.15449	1.14720	1.01310	1.07256	1.11895
17	1.10767	0.89217	0.99410	0.99352	2.28924	0.51559	1.08642	1.07832	0.95895	1.08127	1.06696
18	0.95021	0.83559	0.89105	0.89584	2.14196	0.45252	0.98452	0.97741	0.86911	0.94010	0.94589
19	0.93250	0.81744	0.87308	0.88137	2.21416	0.44435	0.99189	0.98454	0.88768	0.92447	0.93364
20	0.91010	0.85188	0.88051	0.88230	2.37598	0.41411	0.99193	0.98133	0.88109	0.90423	0.92812
21	0.90831	0.87050	0.88920	0.89209	2.48204	0.40411	1.00150	0.99069	0.87548	0.90221	0.92800
22	0.93448	0.79545	0.86217	0.86876	2.44050	0.37816	0.96068	0.95081	0.85191	0.92460	0.90448
23	0.93852	0.82477	0.87981	0.88494	2.54428	0.37672	0.97902	0.96923	0.85916	0.92722	0.86752
24	0.95955	0.83212	0.89357	0.90008	2.61768	0.35461	0.96347	0.95725	0.88900	0.95127	0.87176
25	0.82659	0.77523	0.80050	0.80120	2.54432	0.30555	0.88172	0.87662	0.80638	0.81529	0.78713
26	0.90933	0.75806	0.83026	0.83456	2.84192	0.29847	0.92100	0.91714	0.82419	0.89313	0.85607
27	0.90913	0.86638	0.88749	0.88866	3.22816	0.29960	0.98344	0.97818	0.87350	0.90653	0.87957
28	0.95748	0.71369	0.82665	0.82378	3.27769	0.25120	0.90739	0.90090	0.80609	0.92446	0.88558
29	0.91434	0.79178	0.85086	0.85489	3.58621	0.25612	0.95839	0.95091	0.83824	0.90372	0.92881
30	0.98306	0.74159	0.85383	0.85285	3.63285	0.24640	0.94612	0.93848	0.83636	0.95691	0.90674
31	0.96148	0.79467	0.87411	0.87827	3.82999	0.24849	0.97557	0.96637	0.85604	0.94519	0.95259
32	1.09219	0.77559	0.92038	0.92577	4.36079	0.23020	1.00192	1.00563	0.93404	1.06859	0.93739
33	1.03387	0.82587	0.92403	0.92835	5.45066	0.19325	1.02632	1.03039	0.92860	1.00783	0.98847
34	0.97819	0.92286	0.95012	0.95072	5.95659	0.18655	1.05412	1.05647	0.93004	0.97390	1.01750
35	1.09532	0.90246	0.99422	0.99086	6.44252	0.19130	1.11015	1.11105	0.97904	1.07872	1.09820
36	0.97574	0.93603	0.95568	0.95607	6.69005	0.17668	1.08720	1.08989	0.94745	0.97198	0.93645
37	1.10952	0.99004	1.04808	1.04846	7.50373	0.18937	1.19204	1.19665	1.03628	1.10176	1.02142
38	1.21684	0.99944	1.10280	1.09863	7.90093	0.18768	1.21774	1.22145	1.06914	1.19166	1.14490
39	1.04027	0.86886	0.95071	0.95482	7.16398	0.15715	1.06105	1.06219	0.93030	1.01682	0.99999

Note that the chained Laspeyres index ends up at 7.164 while the chained Paasche index ends up at 0.157. The corresponding fixed base indexes end up at 1.040 and 0.869 so it is clear that these chained indexes are subject to tremendous chain drift. The chain drift carries over to the Fisher and Törnqvist indexes; i.e., the fixed base Fisher index ends up at 0.9548 while its chained counterpart ends up at 1.061. Chart 6 plots the above indexes with the exceptions of the chained Laspeyres and Paasche indexes (these indexes exhibit too much chain drift to be considered further).



It can be seen that all 9 of the weighted indexes which appear on Chart 6 capture an underlying general trend in prices. However, there is a considerable dispersion between the indexes. Our preferred indexes for this group of indexes are the fixed base Fisher and Törnqvist indexes, P_F^t and P_T^t. These two indexes approximate each other very closely and can barely be distinguished in the Chart. The Paasche and Geometric Paasche indexes, P_P^t and P_{GP}^t, lie below our preferred indexes while the remaining indexes generally lie above our preferred indexes. The chained Fisher and Törnqvist indexes, P_{FCh}^t and P_{TCh}^t, approximate each other very closely but both indexes lie well above their fixed base counterparts; i.e., they exhibit a considerable amount of chain drift. Thus chained superlative indexes are not recommended for use with scanner data where the products are subject to large fluctuations in prices and quantities. The fixed base Laspeyres and Geometric Laspeyres indexes, P_L^t and P_{GL}^t, are fairly close to each other and are well above P_F^t and P_T^t. The unit value price index, P_{UV}^t, is subject to large fluctuations and generally lies above our preferred indexes.

We turn now to weighted indexes that use annual weights from a base year.

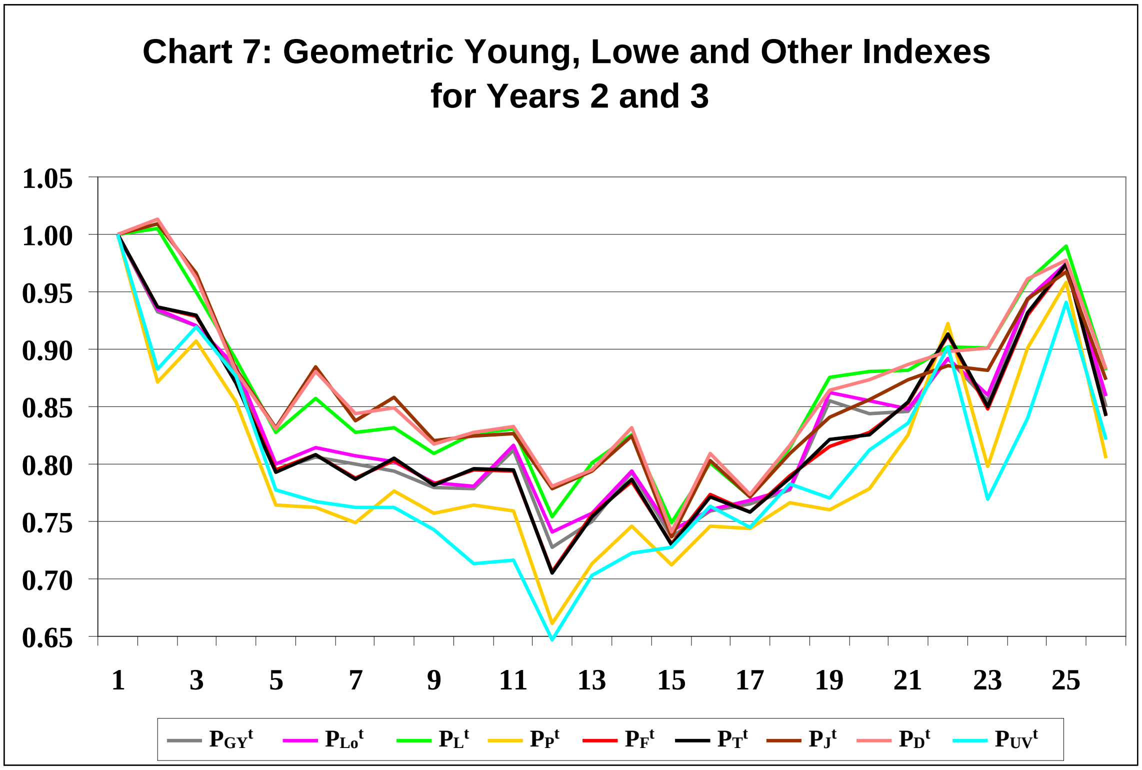
A.4 Indexes which Use Annual Weights

The Weighted Jevons or Geometric Young index, $P_{J\alpha}^t$ or P_{GY}^t, was defined by (58) in section 6. This index uses the arithmetic average of the monthly shares in year 1 as weights in a weighted geometric index for subsequent months in the sample. The Lowe index, P_{Lo}^t, was defined by (124) in section 11. This index is a fixed basket index that uses the average quantities in the base year as the vector of quantity weights. We calculated both of these indexes for the months in years 2 and 3 for our sample using the weights from year 1 of our sample. For comparison purposes, we also list the fixed base Laspeyres, Paasche, Fisher and Törnqvist indexes, P_L^t, P_P^t, P_F^t and P_T^t for the "months" in years 2 and 3 of our sample. It is also of interest to list the Jevons, Dutot and Unit Value indexes, P_J^t, P_D^t and P_{UV}^t for years 2 and 3 in order to see how unweighted indexes compare to the weighted indexes. The sample averages for these indexes are listed in the last row of the table.

Table A.5 Geometric Young, Lowe, Laspeyres, Paasche, Fisher, Törnqvist, Jevons, Dutot and Unit Value Indexes for Years 2 and 3

t	P_{GY}^t	P_{Lo}^t	P_L^t	P_P^t	P_F^t	P_T^t	P_J^t	P_D^t	P_{UV}^t
14	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
15	0.93263	0.93494	1.00555	0.87189	0.93634	0.93715	1.00962	1.01366	0.88282
16	0.92082	0.92031	0.95041	0.90755	0.92873	0.92986	0.96667	0.96175	0.91945
17	0.87997	0.88507	0.89085	0.85384	0.87215	0.87032	0.88275	0.88100	0.87673
18	0.79503	0.80022	0.82733	0.76386	0.79496	0.79283	0.83163	0.83032	0.77724
19	0.80573	0.81468	0.85664	0.76193	0.80790	0.80865	0.88451	0.88091	0.76718
20	0.79981	0.80700	0.82757	0.74904	0.78733	0.78635	0.83739	0.84337	0.76264
21	0.79437	0.80164	0.83126	0.77659	0.80346	0.80489	0.85802	0.84924	0.76254
22	0.77921	0.78355	0.80911	0.75762	0.78294	0.78149	0.82067	0.81723	0.74322
23	0.77876	0.78087	0.82688	0.76468	0.79517	0.79606	0.82440	0.82741	0.71285
24	0.81228	0.81680	0.83070	0.75908	0.79408	0.79472	0.82664	0.83251	0.71633
25	0.72801	0.74112	0.75452	0.66120	0.70632	0.70498	0.77809	0.78023	0.64679
26	0.75011	0.75684	0.80141	0.71350	0.75618	0.75377	0.79362	0.79483	0.70344
27	0.79254	0.79375	0.82527	0.74559	0.78442	0.78661	0.82498	0.83156	0.72275
28	0.73664	0.74226	0.74893	0.71223	0.73035	0.72970	0.73650	0.74098	0.72769
29	0.75964	0.76031	0.80135	0.74576	0.77306	0.77165	0.80321	0.80946	0.76321
30	0.76531	0.76828	0.77149	0.74410	0.75767	0.75781	0.77157	0.77315	0.74507
31	0.77786	0.77867	0.81448	0.76635	0.79005	0.78811	0.80900	0.81594	0.78275
32	0.85506	0.86201	0.87512	0.76018	0.81563	0.82138	0.84067	0.86401	0.77026
33	0.84365	0.85499	0.88099	0.77811	0.82795	0.82554	0.85589	0.87320	0.81223
34	0.84601	0.84804	0.88159	0.82588	0.85328	0.85422	0.87325	0.88632	0.83608
35	0.89199	0.89177	0.90170	0.92254	0.91206	0.91320	0.88570	0.89782	0.90240
36	0.85506	0.85983	0.90132	0.79811	0.84815	0.84966	0.88156	0.90117	0.76948
37	0.94264	0.94402	0.95898	0.90084	0.92946	0.93135	0.94414	0.96137	0.83931
38	0.97419	0.97462	0.99009	0.95811	0.97397	0.97413	0.96684	0.97762	0.94077
39	0.85043	0.85908	0.88213	0.80516	0.84277	0.84144	0.87353	0.88242	0.82170
Mean	0.83338	0.83772	0.86329	0.80014	0.83094	0.83100	0.86080	0.86644	0.79634

As usual, P_F^t and P_T^t approximate each other very closely. Indexes with substantial upward biases relative to these two indexes are the Laspeyres, Jevons and Dutot indexes, P_L^t, P_J^t and P_D^t. The Geometric Young index and the Lowe index, P_{GY}^t and P_{Lo}^t, were about 0.25 and .67 percentage points above the superlative indexes on average. The Paasche and Unit Value indexes, P_P^t and P_{UV}^t, had substantial downward biases relative to the superlative indexes. These inequalities agree with our a priori expectations about biases. The nine indexes are plotted in Chart 7.



It can be seen that all 9 indexes capture the trend in the product prices with P_F^t and P_T^t in the middle of the indexes (and barely distinguishable from each other in the chart). The unit value index P_{UV}^t is the lowest index followed by the Paasche index P_P^t. The Geometric Young and Lowe indexes, P_{GY}^t and P_{Lo}^t, are quite close to each other and close to the superlative indexes in the first part of the sample but then they drift above the superlative indexes in the latter half of the sample. We expect the Lowe index to have some upward substitution bias and with highly substitutable products, we expect the Geometric Young index to also have an upward substitution bias. Finally, the Laspeyres, Jevons and Dutot indexes are all substantially above the superlative indexes with P_J^t and P_D^t approximating each other quite closely.

We turn our attention to multilateral indexes.

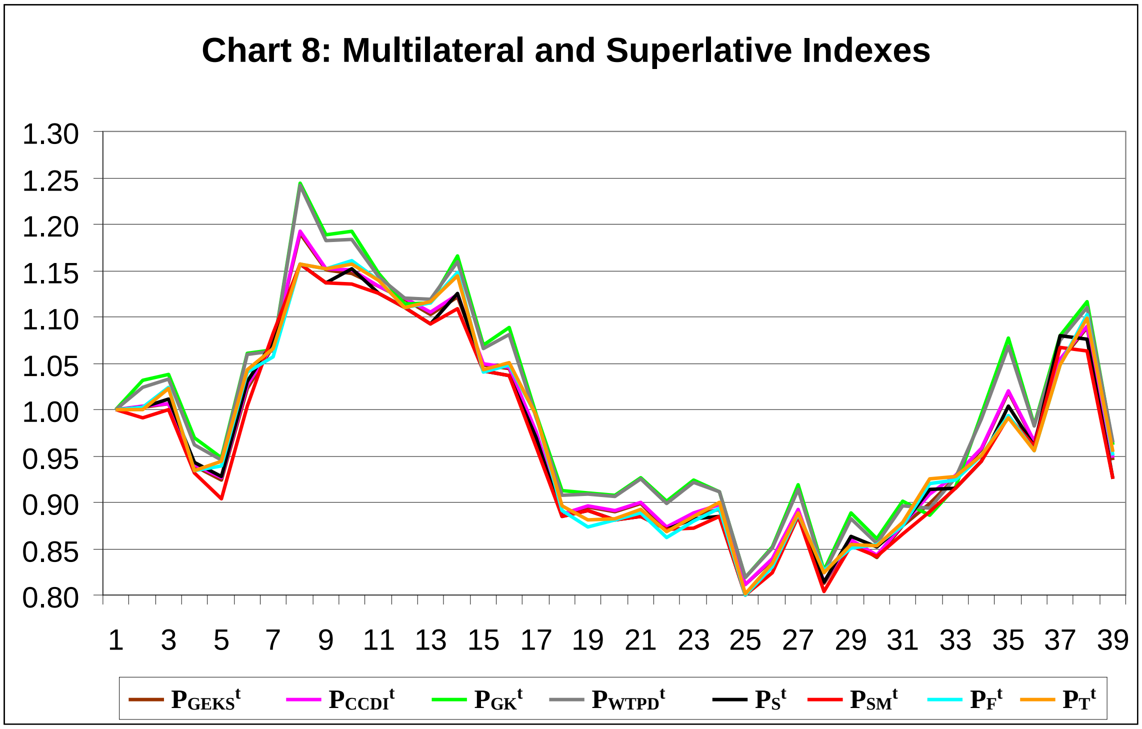
A.5 Multilateral Indexes

We considered 6 multilateral indexes in the main text: P_{GEKS}^t (see definition (70) in section 8); P_{CCDI}^t (see (77) in section 8); P_{GK}^t (see (137) in section 12); P_{WTPD}^t (see (149) in section 13); the real time Similarity linked price indexes P_S^t defined in the beginning of section 14 and the Modified Similarity linked price indexes P_{SM}^t defined in the middle of section 14. These indexes are listed in Table A6 along with the fixed base Fisher and Törnqvist indexes P_F^t and P_T^t. The sample mean for each index is listed in the last row of Table A.6.

Table A.6 Six Multilateral Indexes and the Fixed Base Fisher and Törnqvist Indexes

t	P_{GEKS}^t	P_{CCDI}^t	P_{GK}^t	P_{WTPD}^t	P_S^t	P_{SM}^t	P_F^t	P_T^t
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00233	1.00395	1.03138	1.02468	1.00218	0.99064	1.00218	1.00036
3	1.00575	1.00681	1.03801	1.03322	1.01124	0.99960	1.02342	1.02220
4	0.93922	0.94020	0.97021	0.96241	0.94262	0.93180	0.93388	0.93445
5	0.92448	0.92712	0.94754	0.94505	0.92812	0.90417	0.93964	0.94387
6	1.02249	1.02595	1.06097	1.05893	1.03073	1.00413	1.03989	1.04311
7	1.06833	1.06995	1.06459	1.06390	1.07314	1.08267	1.05662	1.06555
8	1.19023	1.19269	1.24385	1.24192	1.15740	1.15740	1.15740	1.15743
9	1.15115	1.15206	1.18818	1.18231	1.13680	1.13680	1.15164	1.15169
10	1.14730	1.15007	1.19184	1.18333	1.15156	1.13578	1.16081	1.15735
11	1.13270	1.13301	1.14662	1.14308	1.12574	1.12574	1.13876	1.13875
12	1.11903	1.12079	1.11332	1.12082	1.10951	1.10951	1.10951	1.10976
13	1.10247	1.10487	1.11561	1.11838	1.09229	1.09229	1.11511	1.11677
14	1.12136	1.12345	1.16579	1.15912	1.12489	1.10947	1.14803	1.14485
15	1.04827	1.04883	1.06958	1.06608	1.04237	1.04237	1.04086	1.04292
16	1.04385	1.04539	1.08842	1.08044	1.03692	1.03692	1.04836	1.05073
17	0.97470	0.97550	0.99512	0.99145	0.97013	0.95897	0.99410	0.99352
18	0.88586	0.88695	0.91319	0.90765	0.88455	0.88455	0.89105	0.89584
19	0.89497	0.89597	0.90990	0.90923	0.89118	0.89118	0.87308	0.88137
20	0.88973	0.89126	0.90822	0.90578	0.88051	0.88051	0.88051	0.88230
21	0.89904	0.89990	0.92641	0.92503	0.88482	0.88482	0.88920	0.89209
22	0.87061	0.87363	0.90145	0.89880	0.87151	0.87151	0.86217	0.86876
23	0.88592	0.88868	0.92421	0.92158	0.88280	0.87265	0.87981	0.88494
24	0.89282	0.89799	0.91127	0.91198	0.88502	0.88502	0.89357	0.90008
25	0.81132	0.81115	0.81875	0.81913	0.79966	0.79966	0.80050	0.80120
26	0.83799	0.83914	0.85168	0.85089	0.83378	0.82421	0.83026	0.83456
27	0.89063	0.89246	0.91906	0.91398	0.88481	0.88481	0.88749	0.88866
28	0.81304	0.81411	0.82600	0.82419	0.81336	0.80399	0.82665	0.82378
29	0.85763	0.85934	0.88821	0.88248	0.86271	0.85280	0.85086	0.85489
30	0.84103	0.84305	0.86121	0.85556	0.85166	0.84188	0.85383	0.85285
31	0.87495	0.87639	0.90123	0.89600	0.87568	0.86562	0.87411	0.87827
32	0.89936	0.90831	0.88553	0.89332	0.91368	0.89010	0.92038	0.92577
33	0.92670	0.92878	0.91672	0.92625	0.91517	0.91517	0.92403	0.92835
34	0.95721	0.95846	0.99507	0.98974	0.94435	0.94435	0.95012	0.95072
35	1.01848	1.02026	1.07728	1.06779	1.00422	0.99266	0.99422	0.99086
36	0.96507	0.96601	0.98339	0.98282	0.96122	0.96122	0.95568	0.95607
37	1.05250	1.05448	1.08019	1.07514	1.07953	1.06710	1.04808	1.04846
38	1.08819	1.08961	1.11648	1.10963	1.07546	1.06308	1.10280	1.09863
39	0.94591	0.94834	0.96156	0.96453	0.92575	0.92575	0.95071	0.95482
Mean	0.97417	0.97602	0.99764	0.99504	0.97069	0.96464	0.97434	0.97607

If the 8 indexes are evaluated according to their sample means, the Modified Similarity Linked index P_{SM}^t generates the lowest indexes followed by the real time Similarity index P_S^t. The P_{GEKS}^t, P_{GEKS}^t, P_F^t and P_T^t indexes are tightly clustered in the middle and the P_{GK}^t and P_{WTPD}^t are about 2 percentage points above the middle indexes on average. Chart 8 plots the 8 indexes.



All 8 indexes capture the trend in product prices quite well. It is clear that the Geary-Khamis and Weighted Time Product Dummy indexes have a substantial upward bias relative to the remaining 6 indexes. Although P_S^t and P_{SM}^t both end up at the same index level, the Modified Similarity linked indexes are on average 0.6 percentage points below the corresponding Similarity linked indexes. P_S^t and P_{SM}^t end up about 2 percentage points below P_{GEKS}^{39} and P_{CCDI}^{39} and almost 4 percentage points below P_{GK}^{39} and P_{WTPD}^{39}. Thus there are significant differences between the various multilateral indexes.

Conceptually, the Modified Similarity linked indexes seem to be the most attractive solution for solving the chain drift problem.[132]

A.6 Multilateral Indexes Using Inflation Adjusted Carry Forward and Backward Prices

In section 15, we used inflation adjusted carry forward and carry backward prices to replace the reservation prices for the missing products. In this section, we recalculate the 8 indexes listed on Table A.6 above where the econometrically estimated reservation prices are replaced by these carry forward and backward prices.[133] Our hope is that the

[132] They can deal with seasonal products more adequately than the other indexes that are considered in this paper.

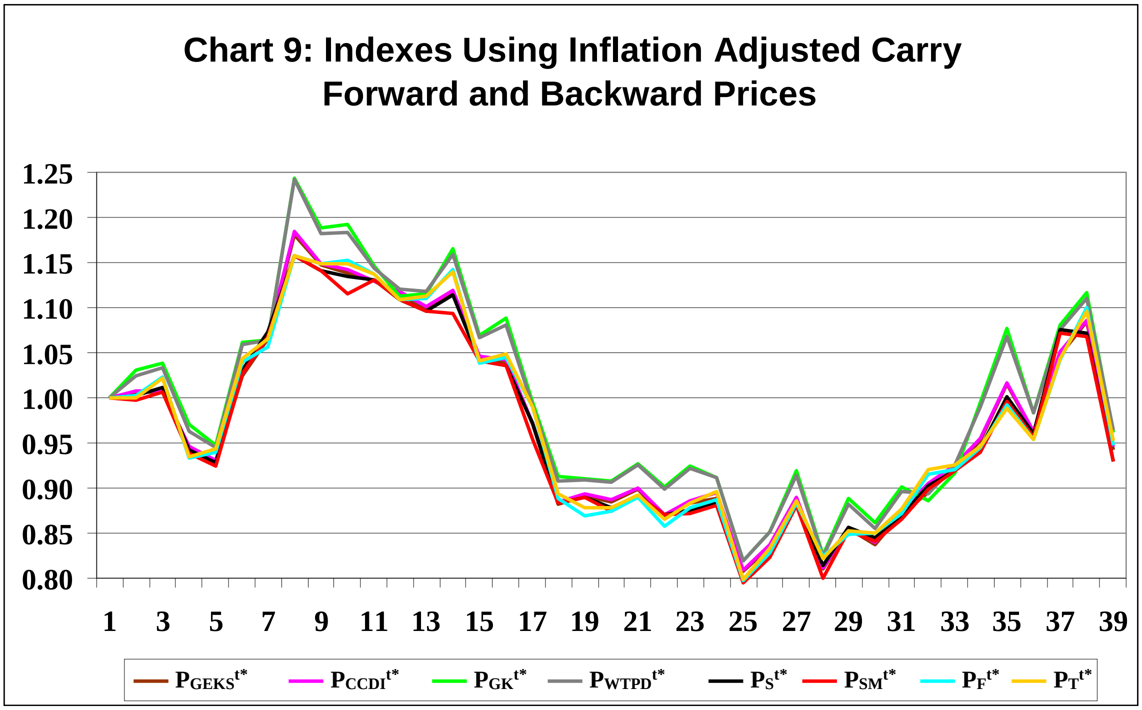
[133] The 8 inflation adjusted carry backward prices for products 2 and 4 are: 0.14717 0.15011 0.14755 0.14154 0.13754 0.15373 0.17102 0.16925 and 0.14115 0.14397 0.14151 0.13575 0.13191 0.14744 0.16401 0.16232 respectively. The inflation adjusted carry forward prices for product 12 are: 0.07399 0.08811 0.08854 0.08460.

new indexes will be close to the previously calculated indexes in which case it will not be necessary for statistical agencies to expend resources in order to calculate reservation prices. The new indexes are differentiated from the old indexes by adding asterisks; e.g., the new Fisher index for period t is denoted as P_F^{t*}, etc. The sample means for the new indexes are listed in the last row of Table A.7.

Table A.7 Multilateral Indexes and the Fisher and Törnqvist Indexes Using Inflation Adjusted Carry Forward and Backward Prices

t	P_{GEKS}^{t*}	P_{CCDI}^{t*}	P_{GK}^{t*}	P_{WTPD}^{t*}	P_S^{t*}	P_{SM}^{t*}	P_F^{t*}	P_T^{t*}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00596	1.00730	1.03138	1.02468	1.00218	0.99792	1.00218	1.00036
3	1.00666	1.00759	1.03801	1.03322	1.01124	1.00691	1.02342	1.02220
4	0.94647	0.94621	0.97021	0.96241	0.94262	0.93862	0.93388	0.93445
5	0.92870	0.93022	0.94754	0.94505	0.92812	0.92415	0.93964	0.94387
6	1.02388	1.02738	1.06097	1.05893	1.03073	1.02632	1.03989	1.04311
7	1.06561	1.06763	1.06459	1.06390	1.07314	1.06858	1.05662	1.06555
8	1.18091	1.18406	1.24385	1.24192	1.15740	1.15740	1.15740	1.15743
9	1.14713	1.14861	1.18818	1.18231	1.14132	1.14132	1.14877	1.14899
10	1.13877	1.14192	1.19184	1.18333	1.13504	1.11475	1.15261	1.14931
11	1.12911	1.13001	1.14662	1.14308	1.13022	1.13022	1.13709	1.13731
12	1.11560	1.11787	1.11332	1.12082	1.10844	1.10844	1.10844	1.10868
13	1.09803	1.10122	1.11561	1.11838	1.09664	1.09664	1.11002	1.11282
14	1.11652	1.11948	1.16579	1.15912	1.11362	1.09370	1.14194	1.13993
15	1.04479	1.04585	1.06958	1.06608	1.04137	1.04137	1.03879	1.04091
16	1.04010	1.04230	1.08842	1.08044	1.03592	1.03592	1.04526	1.04821
17	0.97090	0.97242	0.99512	0.99145	0.97124	0.95387	0.99011	0.99035
18	0.88258	0.88427	0.91319	0.90765	0.88370	0.88370	0.88815	0.89350
19	0.89145	0.89306	0.90990	0.90923	0.89032	0.89032	0.86971	0.87847
20	0.88485	0.88728	0.90822	0.90578	0.87840	0.87500	0.87500	0.87811
21	0.89905	0.90051	0.92641	0.92503	0.89051	0.89051	0.88918	0.89266
22	0.86702	0.87076	0.90145	0.89880	0.87100	0.87100	0.85782	0.86544
23	0.88301	0.88622	0.92421	0.92158	0.87581	0.87209	0.87827	0.88348
24	0.88841	0.89459	0.91127	0.91198	0.88388	0.88046	0.88688	0.89553
25	0.80827	0.80863	0.81875	0.81913	0.79839	0.79530	0.79772	0.79902
26	0.83511	0.83675	0.85168	0.85089	0.82719	0.82367	0.82832	0.83286
27	0.88704	0.88948	0.91906	0.91398	0.88380	0.88038	0.88358	0.88553
28	0.81023	0.81176	0.82600	0.82419	0.81428	0.79972	0.82477	0.82203
29	0.85467	0.85682	0.88821	0.88248	0.85588	0.85224	0.84893	0.85297
30	0.83762	0.84022	0.86121	0.85556	0.84492	0.84133	0.85028	0.84975
31	0.87196	0.87386	0.90123	0.89600	0.86875	0.86505	0.87220	0.87638
32	0.89520	0.90478	0.88553	0.89332	0.90195	0.89810	0.91492	0.92083
33	0.92324	0.92594	0.91672	0.92625	0.91881	0.91881	0.92070	0.92557
34	0.95312	0.95505	0.99507	0.98974	0.94327	0.93963	0.94499	0.94645
35	1.01483	1.01725	1.07728	1.06779	1.00095	0.99667	0.99168	0.98858
36	0.96184	0.96328	0.98339	0.98282	0.96013	0.95642	0.95363	0.95423
37	1.04792	1.05071	1.08019	1.07514	1.07601	1.07141	1.04224	1.04379
38	1.08429	1.08637	1.11648	1.10963	1.07195	1.06737	1.09941	1.09567
39	0.94251	0.94552	0.96156	0.96453	0.92943	0.92943	0.94795	0.95253
Mean	0.97137	0.97367	0.99764	0.99504	0.96894	0.96499	0.97160	0.97377

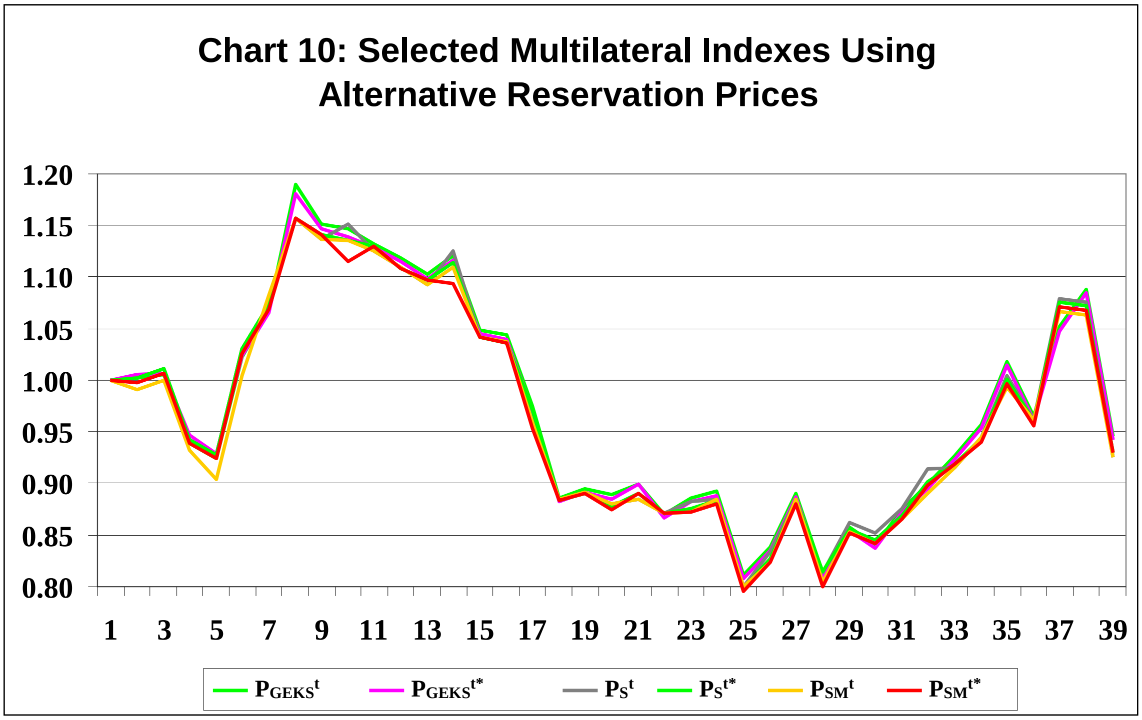
The new indexes using carried prices are plotted on Chart 9. A comparison of Charts 8 and 9 shows that there is little change in the 8 indexes when inflation adjusted carry forward or backward prices replace the econometrically estimated reservation prices for the missing products.



It can be seen that there is very little difference between Charts 8 and 9: The P_{GK}^{t*} and P_{WTPD}^{t*} price indexes using carried prices are still well above the rest of the indexes and the Modified Similarity linked indexes, P_{SM}^{t*}, are still below the other indexes for most observations.

If the means listed in Tables A.6 and A.7 are compared, then it can be seen that since the P_{GK}^{t} and P_{WTPD}^{t} do not depend on reservation prices, the means for these indexes remain unchanged. The means for P_{GEKS}^{t}, P_{CCDI}^{t}, P_{F}^{t} and P_{T}^{t} all decrease by about 0.3 percentage points when carried prices are substituted for reservation prices. The mean for P_{S}^{t} decreased by 0.18 percentage points when the reservation prices were replaced by carried prices. However, the mean for the Modified Similarity linked prices increased by only 0.035 percentage points when the reservation prices were replaced by carried prices. Thus the substitution of inflation adjusted carry forward and backward prices hardly affected the Modified Similarity linked indexes. This is very encouraging result. If this result were to hold more widely, then the calculation of similarity linked price indexes is an attractive practical alternative that can deal adequately with both seasonal commodities and the chain drift problem .

Chart 10 plots P_{GEKS}^t, P_S^t and P_{SM}^t with the counterpart indexes that used inflation adjusted carry forward and backward prices for the missing products, P_{GEKS}^{t*}, P_S^{t*} and P_{SM}^{t*}.



It can be seen that P_{GEKS}^t closely approximates P_{GEKS}^{t*} and both of these indexes generally lie above the similarity linked indexes. The similarity linked indexes are tightly bunched with P_{SM}^{t*} dipping below the other indexes occasionally.

References

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

.

Alterman, W.F.., W.E. Diewert and R.C. Feenstra (1999), *International Trade Price Indexes and Seasonal Commodities*, Bureau of Labor Statistics, Washington D.C.

Armknecht, P. and M. Silver (2014), "Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indexes", *Review of Income and Wealth* 60:2, 225-244.

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.

Aten, B. and A. Heston (2009), "Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives", pp. 245-273 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

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

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

Balk, B.M. (1981), "A Simple Method for Constructing Price Indices for Seasonal Commodities", *Statistische Hefte* 22 (1), 1–8.

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. (2008), *Price and Quantity Index Numbers*, New York: Cambridge University Press.

Bortkiewicz, L.v. (1923), "Zweck und Struktur einer Preisindexzahl", *Nordisk Statistisk Tidsskrift* 2, 369-408.

Carli, Gian-Rinaldo, (1804), "Del valore e della proporzione de' metalli monetati", pp. 297-366 in *Scrittori classici italiani di economia politica*, Volume 13, Milano: G.G. Destefanis (originally published in 1764).

Carruthers, A.G., D.J. Sellwood and P.W. Ward (1980), "Recent Developments in the Retail Prices Index", *The Statistician* 29, 1-32.

Caves D.W., Christensen, L.R. and Diewert, W.E. (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* 1/2016, 49-69.

Cobb, Charles W. and Paul H. Douglas (1928), "A Theory of Production," *American Economic Review* 18(1): 139-165.

Dalén, J. (1992), "Computing Elementary Aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics* 8, 129-147.

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

Davies, G.R. (1924), "The Problem of a Standard Index Number Formula", *Journal of the American Statistical Association* 19, 180-188.

Davies, G.R. (1932), "Index Numbers in Mathematical Economics", *Journal of the American Statistical Association* 27, 58-64.

de Haan, J. (2004), "Estimating Quality-Adjusted Unit Value Indexes: Evidence from Scanner Data," Paper presented at the SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity, June 30-July 3, 2004, Vancouver.

de Haan, J. (2015b), "Rolling Year Time Dummy Indexes and the Choice of Splicing Method", Room Document at the 14th meeting of the Ottawa Group, May 22, Tokyo. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s3room

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*, forthcoming.

Diewert, W.E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", Discussion Paper No. 95-01, Department of Economics, University of British Columbia, Vancouver, Canada.

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

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

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

Diewert, W.E. (1999), "Axiomatic and Economic Approaches to International Comparisons", pp. 13-87 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Volume 61, Chicago: The University of Chicago Press.

Diewert, W.E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Department of Economics, Discussion Paper 04-16, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Diewert, W.E. (2005), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", *The Review of Income and Wealth* 51:4, 561-571.

Diewert, W.E. (2009), "Similarity Indexes and Criteria for Spatial Linking", pp. 183-216 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham, UK: Edward Elgar.

Diewert, W.E. (2012), *Consumer Price Statistics in the UK*, Government Buildings, Cardiff Road, Newport, UK, NP10 8XG: Office for National Statistics. http://www.ons.gov.uk/ons/guide-method/userguidance/prices/cpi-and-rpi/index.html

Diewert, W.E. (2013), "Methods of Aggregation above the Basic Heading Level within Regions", pp. 121-167 in *Measuring the Real Size of the World Economy: The Framework, Methodology and Results of the International Comparison Program—ICP*, Washington D.C.: The World Bank.

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

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

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

Diewert, W.E. and P. von der Lippe (2010), "Notes on Unit Value Index Bias", *Journal of Economics and Statistics* 230, 690-708.

Drobisch, M.W. (1871), "Über die Berechnung der Veränderung der Waarenpreis und des Geldwertes", *Jahrbücher für Nationalökonomie und Statistik* 16, 416-427.

Dutot, Charles, (1738), *Réflexions politiques sur les finances et le commerce*, Volume 1, La Haye: Les frères Vaillant et N. Prevost.

Eltetö, Ö., and Köves, P. (1964), "On a Problem of Index Number Computation Relating to International Comparisons", (in Hungarian), *Statisztikai Szemle* 42, 507-518.

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

Feenstra, R.C. and M.D. Shapiro (2003), "High-Frequency Substitution and the Measurement of Price Indexes", pp. 123-146 in *Scanner Data and Price Indexes*, Robert C. Feenstra and Matthew D. Shapiro (eds.), Studies in Income and Wealth Volume 64, Chicago: The University of Chicago Press.

Fisher, I. (1911), *The Purchasing Power of Money*, London: Macmillan.

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

Frisch, R. (1936), "Annual Survey of General Economic Theory: The Problem of Index Numbers", *Econometrica* 4, 1-39.

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.

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

Hicks, J.R. (1940), "The Valuation of the Social Income", *Economica* 7, 105-140.

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. (1999a), "Comparing Price Levels across Countries Using Minimum Spanning Trees", *The Review of Economics and Statistics* 81, 135-142.

Hill, R.J. (1999b), "International Comparisons using Spanning Trees", pp. 109-120 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth Volume 61, NBER, Chicago: The University of Chicago Press.

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.

Hill, R.J. (2009), "Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies", pp. 217-244 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Hill, R.J., D.S. Prasada Rao, S. Shankar and R. Hajargasht (2017), "Spatial Chaining as a Way of Improving International Comparisons of Prices and Real Incomes", paper presented at the Meeting on the International Comparisons of Income, Prices and Production, Princeton University, May 25-26.

Hill, T.P. (1988), "Recent Developments in Index Number Theory and Practice", *OECD Economic Studies* 10, 123-148.

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

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, Vancouver, Canada.

Ivancic, L., W.E. Diewert and K.J. Fox (2010), "Using a Constant Elasticity of Substitution Index to estimate a Cost of Living Index: from Theory to Practice", Australian School of Business Research Paper No. 2010 ECON 15, University of New South Wales, Sydney 2052 Australia

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.

Khamis, S.H. (1972), "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society Series A* 135, 96-121.

Kravis, I.B., A. Heston and R. Summers (1982), *World Product and Income: International Comparisons of Real Gross Product*, Statistical Office of the United Nations and the World Bank, Baltimore: The Johns Hopkins University Press.

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

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

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

Konüs, A.A. and S.S. Byushgens (1926), “K probleme pokupatelnoi cili deneg”, *Voprosi Konyunkturi* 2, 151-172.

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

Laspeyres, E. (1871), “Die Berechnung einer mittleren Waarenpreissteigerung”, *Jahrbücher für Nationalökonomie und Statistik* 16, 296-314.

Leontief, W. (1936), “Composite Commodities and the Problem of Index Numbers”, *Econometrica* 4, 39-59.

Lowe, J. (1823), *The Present State of England in Regard to Agriculture, Trade and Finance,* second edition, London: Longman, Hurst, Rees, Orme and Brown.

Marris, R. (1984), “Comparing the Incomes of Nations: A Critique of the International Comparison Project”, *Journal of Economic Literature* 22:1, 40-57.

Paasche, H. (1874), “Über die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen”, *Jahrbücher für Nationalökonomie und Statistik* 12, 168-178.

Persons, W.M. (1921), “Fisher's Formula for Index Numbers”, *Review of Economics and Statistics* 3:5, 103-113.

Persons, W.M. (1928), “The Effect of Correlation Between Weights and Relatives in the Construction of Index Numbers”, *The Review of Economics and Statistics* 10:2, 80-107.

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

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

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

Schlömilch, O., (1858), “Über Mïttelgrössen verschiedener Ordnungen”, *Zeitschrift für Mathematik und Physik* 3, 308-310.

Sergeev, S. (2001), “Measures of the Similarity of the Country’s Price Structures and their Practical Application”, Conference on the European Comparison Program, U. N. Statistical Commission. Economic Commission for Europe, Geneva, November 12-14, 2001.

Sergeev, S. (2009), “Aggregation Methods Based on Structural International Prices”, pp. 274-297 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Shapiro, M.D. and D.W. Wilcox (1997), “Alternative Strategies for Aggregating Prices in the CPI”, *Federal Reserve Bank of St. Louis Review* 79:3, 113-125.

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

Szulc, B.J. (1964), “Indices for Multiregional Comparisons”, (in Polish), *Przeglad Statystyczny* 3, 239-254.

Szulc, B.J. (1983), “Linking Price Index Numbers,” pp. 537-566 *in Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Szulc, B.J. (1987), “Price Indices below the Basic Aggregation Level”, *Bulletin of Labour Statistics* 2, 9-16.

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

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

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

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

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

Vartia, Y.O. (1978), "Fisher's Five-Tined Fork and Other Quantum Theories of Index Numbers, pp. 271-295 in *Theory and Applications of Economic Indices*, W. Eichhorn, R. Henn, O. Opitz, R. W. Shephard (eds.), Wurzburg: Physica-Verlag.

Vartia, Y. and A. Suoperä (2018), "Contingently Biased, Permanently Biased and Excellent Index Numbers for Complete Micro Data", unpublished paper: http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/contingently_biased_vartia_suopera_updated.pdf

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

Walsh, C. M. (1921), "Discussion", *Journal of the American Statistical Association* 17, 537-544.

Young, A. (1812), *An Inquiry into the Progressive Value of Money as Marked by the Price of Agricultural Products*, London: McMillan.