Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data

W. Erwin Diewert

University of British Columbia and UNSW Sydney

Kevin J. Fox[*]

UNSW Sydney

October 2018

Abstract

The use of multilateral indexes is increasingly an accepted approach for incorporating scanner data in a Consumer Price Index. The attractiveness stems from the ability to be able to control for chain drift bias. Consensus on two key issues has yet to be achieved: (i) the best multilateral method to use, and (ii) the best way of extending the resulting series when new observations become available. We present theoretical and simulation evidence on the extent of substitution biases in alternative methods. Our results suggest the use of the CCDI index with a new method, the "mean splice", for updating.

Keywords:Consumer Price Index, chain drift, multilateral indexes, Rolling Window indexes, linking methods.
JEL Classification Numbers: C43, E31

1   Introduction

Several national statistical agencies are either using or considering the use of scanner data from retail outlets that collect data on weekly unit value prices and the corresponding quantities sold of specific products. These detailed data on prices and quantities can be used in order to construct consumer price indexes for various product groups. Besides the benefits from using more information in a more efficient way, there are additional benefits that can arise. These include the potential for improving the timeliness of CPI releases and the potential to move resources from data collection and processing to analytics.[1]

The 2004 International Labour Office *Consumer Price Index Manual* recommended that chained superlative indexes be used in constructing national Consumer Price Indexes (CPIs) when detailed price and quantity data are available. The reason why chaining was recommended over the use of fixed base indexes was that more exact product matches across periods will generally be found if chained indexes are constructed instead of fixed base indexes; this is due to the fact that often one or two percent of products disappear from the marketplace each month in many countries.[2] However, as more experience with the use of chained indexes has become available, it has been noticed that chained indexes frequently drift downwards as compared to corresponding fixed base indexes. It is a serious problem and this paper examines possible solutions.

CPIs have traditionally been constructed by statistical agencies using *bilateral index number theory*; i.e., the prices and quantities for a group of commodities for the current period were compared to the same prices and quantities for a base period. *Multilateral index number theory* uses the prices and quantities for a group of commodities for multiple periods (say T periods) and simultaneously constructs a sequence of price indexes for all T periods. Ivancic, Diewert and Fox (2011) suggested that the use of multilateral indexes in the scanner data context can largely solve the chain drift problem. Subsequently, there has been a considerable amount of additional research on the use of multilateral methods

---

[1]The Australian Bureau of Statistics implemented the approach recommended in this paper in CPI construction, starting from the December Quarter of 2017; see the Australian Bureau of Statistics (2017)(2018).

[2]Using Adobe Analytics data on online transactions from January 2014 to September 2017, Goolsbee and Klenow (2018; 490) found that "roughly half of the sales volume online is for products that did not exist in the previous year. Even without apparel, the figure is 44%. The products that disappear, meanwhile, had about 24% of total sales before they left the market (22% excluding apparel)."

when detailed price and quantity data are available. We review and extend this research.

The chain drift problem is explained in more detail in section 2. Multilateral index methods are explained in section 3, starting with the GEKS multilateral method due to Gini (1931; 12), Eltetö and Köves (1964) and Szulc (1964). The GEKS method uses the bilateral Fisher (1922) ideal index as a basic building block. Instead of using the Fisher index, it is possible to use Gini's basic methodology but use the Törnqvist index as the bilateral building block index. This idea was used by Caves, Christensen and Diewert (1982) using distance functions, and adapted to the price index context by Inklaar and Diewert (2016). This multilateral method, which we call the CCDI approach, will be explained in section 3.2.[3] Section 3.3 reviews and introduces some new results on the Weighted Time Product Dummy (WTPD) method which is a generalization of Summer's (1973) Country Product Dummy method for making price comparisons across countries. Section 3.4 examines the Geary (1958) Khamis (1970) (1972) (GK) multilateral method.[4]

A multilateral index number method will determine a sequence of price indexes or aggregate price levels for say T periods where $T \geq 3$. When the data for period $T+1$ becomes available, CPI methodology typically does not allow for a revision of price levels that have already been determined for periods 1 to T, so those initially determined price levels remain the same. If the multilateral method is used to form a new sequence of price levels for periods 2 to $T+1$, the following question arises: how exactly should the new price levels be linked to the initial sequence of price levels? Various methods for accomplishing this linking have been suggest in the recent literature, and in section 3.5 we will study these alternative linking methods in the context of the CCDI multilateral indexes since this facilitates comparisons.

Section 4 introduces Constant Elasticity of Substitution (CES) purchaser preferences as a basis for a target cost of living index.[5] Although attracted to multilateral methods for

---

[3]Given that it is a variant of the GEKS approach advocated by Ivancic, Diewert and Fox (2011) in the scanner data context, it is sometimes called "GEKS-Törnqvist.

[4]Ivancic, Diewert and Fox (2009) (2011) used the GEKS and the Weighted Time Product Dummy multilateral methods which will be explained in section 3. Khamis (1970; 83-85) (1972; 101) noted that Geary-Khamis multilateral indexes could be applied in the time series context as well as in the international comparisons context. Balk (1981) also adapted Gini's basic methodology to the time series context using the Sato (1976) Vartia (1976) bilateral index number formula as the bilateral building block formula.

[5]Balk's (1981) multilateral method, which adapted Gini's basic methodology using the Sato-Vartia bilateral price index as the basic building block, is exact for CES preferences.

using scanner data in price indexes, statistical agencies have had difficulties in trying to determine which multilateral method should be used to aggregate price data when detailed price and quantity data are available for a stratum of their CPIs. The problem is that the "truth" is not known. Thus in the later sections of this study, we will assume that we know the "truth"; i.e., we will assume that purchasers have known CES preferences and thus we can construct the corresponding true cost of living indexes given an artificial data set. Then the various multilateral methods that have been suggested in the literature can be constructed using the artificial data set and compared with the corresponding true CES cost of living indexes. The assumption of CES preferences has been used in many economics and marketing studies.[6]

We present our first artificial data set in section 4.1. We assume given price data on four commodities for twelve periods along with total expenditures on the commodities. CES preferences are assumed but we allow the elasticity of substitution parameter σ to take on six separate values: 0, 0.5, 1, 2, 4 and 10. For each of these values of σ, we construct quantity data based on the CES functional form. As σ varies, the price data remain the same but the quantity data change in order to be consistent with the CES preferences. In section 4.1, we construct the six variants of the artificial data set along with the true CES cost of living indexes and compare these true indexes with fixed base and chained Fisher and Törnqvist indexes as well as the single window GK, WTPD, GEKS and CCDI multilateral price levels. It turns out that the best approximations to the true indexes vary as the elasticity of substitution varies.

The methodological approach taken in section 4.1 does not address the chain drift problem because it does not arise using the artificial data set. Hence in section 4.2, we alter the data used in section 4.1 to reflect the chain drift problem. This problem is basically caused by heavily discounted prices. Purchasers stock up on a product when it goes on sale. In the period following the sale, consumers do not purchase the same amount of the product that they did in the pre-sale period. This behaviour is not consistent with the assumptions underlying the typical theory of the cost of living index, which does not

---

[6]For economics studies, see e.g. Balk (1981), Feenstra (1994), Ivancic, Diewert and Fox (2010), and Goolsbee and Klenow (2018), and Baltas (2001) for an example from the marketing literature.

take into account the holding of stocks of consumer goods over multiple periods.[7] The net effect of this stockpiling behaviour is to cause chained indexes to have a downward bias as compared to a corresponding fixed base index.

In section 4.3, we use the adjusted artificial data set constructed in section 4.2 to study the window linking problem of section 3.5. In section 4.4, we again use the adjusted artificial data set in order to introduce two additional multilateral methods that are based on linking observations that have the most similar structure of relative prices. Section 5 concludes.

2 The Chain Drift Problem

More than a decade has passed since the 2004 *Consumer Price Index Manual* was published.[8] The main problem that the *Manual* did not address adequately is the *chain drift*.[9] Suppose that a chosen group of households purchases N quantities in each of two periods and the representative basket is defined by the positive quantity vector $q \equiv [q_1, \ldots, q_N]$. Given the price vectors for periods 1 and 2, $p^1 \equiv [p_1^1, \ldots, p_N^1]$ and $p^2 \equiv [p_1^2, \ldots, p_N^2]$ respectively, we can calculate the cost of purchasing this same basket in the two periods, $p^1 \cdot q \equiv \sum_{n=1}^{N} p_n^1 q_n$ and $p^2 \cdot q \equiv \sum_{n=1}^{N} p_n^2 q_n$. This leads to the *Lowe* (1823) *price index*, P_{Lo}, which is commonly used in CPI construction, defined as follows:

$$P_{Lo}(p^1, p^2, q) \equiv p^1 \cdot q / p^1 \cdot q. \tag{1}$$

There are two natural choices for the reference basket: the period 1 commodity vector q^1 that was actually purchased by the group of households or the corresponding period 2 commodity vector q^2.[10] These two choices lead to the Laspeyres (1871) price index P_L

---

[7]In principle, this stockpiling activity could be modeled by treating the purchases as purchases of durable goods and constructing user costs.

[8]See the ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004). The Manual was written over the years 2000-2003. For brevity, in the future, we will refer to the CPI Manual as the *Manual*.

[9]Szulc (1983) (1987) demonstrated how big the chain drift problem could be with chained Laspeyres indexes but the authors of the Manual did not realize that chain drift could also be a problem with chained superlative indexes.

[10]In practice, the reference basket is often determined by the last survey of household expenditures.

defined by (2) and the Paasche (1874) price index P_P defined by (3):

$$P_L(p^1, p^2, q^1, q^2) \equiv p^2 q^1 / p^2 q^1 = \sum_{n=1}^{N} s_n^1 (p_n^2 / p_n^1); \tag{2}$$

$$P_P(p^1, p^2, q^1, q^2) \equiv p^2 q^2 / p^1 q^2 = \left[\sum_{n=1}^{N} s_n^2 (p_n^2 / p_n^1)^{-1} \right]^{-1} \tag{3}$$

where the period t expenditure share on commodity n, s_n^t, is defined as $p_n^t q_n^t / p^t q^t$ for $n = 1, \ldots, N$ and $t = 1, 2$. Thus the Laspeyres price index P_L can be written as a base period expenditure share weighted average of the N price ratios (or price relatives), p_n^2 / p_n^1. The last equation in (3) shows that the Paasche price index P_P can be written as a period 2 (or current period) expenditure share weighted *harmonic* average of the N price ratios.

A problem with the Lowe, Laspeyres and Paasche indexes is that they are subject to *substitution bias*. Typically, when the price of a commodity decreases, consumers purchase more of it and conversely when a price increases, consumers purchase less of it. Diewert (1976) introduced the concept of a *superlative index number formula*, which deals with this problem of substitution bias.[11] Fisher (1922) suggested taking the geometric mean of the Paasche and Laspeyres indexes, which results in a superlative index, P_F:

$$P_F(p^1, p^2, q^1, q^2) \equiv [P_L(p^1, p^2, q^1, q^2) P_P(p^1, p^2, q^1, q^2)]^{1/2}. \tag{4}$$

Another commonly used superlative index is the Törnqvist index, P_T:[12]

$$P_T(p^1, p^2, q^1, q^2) \equiv \prod_{n=1}^{N} \left(p_n^2 / p_n^1 \right)^{1/2(s_n^1 + s_n^2)}. \tag{5}$$

---

[11] A superlative index number formula has the property that it is exactly equal to a Konüs (1924) true cost of living index provided that the purchasing households have preferences that can be represented by certain functional forms, where these functional forms can approximate arbitrary preferences to the accuracy of a second order approximation.

[12] This index first appeared explicitly as formula 123 in Fisher (1922; 473). P_T is generally attributed to Törnqvist (1936) but this article did not have an explicit definition for P_T; it was defined explicitly in Törnqvist and Törnqvist (1937); see Balk (2008; 26). This index requires that all prices in each period are positive, which is a restrictive assumption. Implicitly, we assume that if say commodity n is not available in period t so that $q_n^t = 0$, then there is a positive reservation price $p_n^t > 0$ that will induce potential purchasers to demand a zero amount of the commodity. This methodological approach to new and disappearing goods follows Hicks (1940; 114). The practical problem facing price statisticians is: how exactly are these reservation prices to be determined? See e.g. Diewert and Feenstra (2017) and Brynjolfsson, Collis, Diewert, Eggers and Fox (2018) for possible solutions.

The Fisher and Törnqvist indexes were recommended in the *Manual* as target indexes.[13]

In order to explain the chain drift problem, it is first necessary to explain how chained indexes differ from fixed base indexes. Suppose that we have decided on a "best" price index formula that compares the prices of period 1 with those of period 2, say $P(p^1, p^2, q^1, q^2)$. Suppose further that we have price and quantity data for three periods. The sequence of the price levels for the three periods under consideration, P^1, P^2 and P^3, using *fixed base* (or direct) *indexes* can be constructed as follows:

$$P^1 \equiv 1; \quad P^2 \equiv P(p^1, p^2, q^1, q^2); \quad P^3 \equiv P(p^1, p^3, q^1, q^3). \tag{6}$$

Thus the prices in period 3, p^3, are compared *directly* with the prices in period 1, p^1.

The sequence of the three price levels, P^1, P^2 and P^3, using *chained indexes* can be constructed as follows:

$$P^1 \equiv 1; \quad P^2 \equiv P(p^1, p^2, q^1, q^2); \quad P^3 \equiv P(p^1, p^2, q^1, q^2)P(p^2, p^3, q^2, q^3). \tag{7}$$

Thus fixed base and chained price levels coincide for the first two periods but in subsequent periods t, the fixed base indexes compare the prices in period t directly to the prices in period 1 whereas the chained indexes simply update the price level in the previous period by multiplying by the period over period *chain link index* $P(p^{t-1}, p^t, q^{t-11}, q^t)$. The two methods of index construction will coincide if the bilateral price index formula $P(p^1, p^2, q^1, q^2)$ satisfies the following test:

$$\textit{Circularity Test} : P(p^1, p^2, q^1, q^2)P(p^2, p^3, q^2, q^3) = P(p^1, p^3, q^1, q^3). \tag{8}$$

If there is only one commodity in the aggregate, then the price index $P(p^1, p^2, q^1, q^2)$ just becomes the single price ratio, p_1^2/p_1^1, and the circularity test becomes the equation $[p_1^2/p_1^1][p_1^3/p_1^2] = [p_1^3/p_1^1]$, which is obviously satisfied. The equation in the circularity test

---

[13]In the Manual, these indexes were regarded as preferred not only from the viewpoint of the economic approach to index number theory but also from the viewpoint of the test approach (for the Fisher index) and from the viewpoint of the stochastic approach (for the Törnqvist index; see Theil (1967; 136-137)). The US Bureau of Labor Statistics uses the Törnqvist price index as its target index for its chained Consumer Price Index (CPI); see Bureau of Labor Statistics (2007).

illustrates the difference between chained index numbers and fixed base index numbers. The left hand side of (8) uses the *chain principle* to construct the overall inflation between periods 1 and 3 whereas the right hand side uses the *fixed base principle* to construct an estimate of the overall price change between periods 1 and 3.[14]

It would be ideal if our preferred index number formulae, the Fisher and Törnqvist indexes (P_F and P_T), satisfied the circularity test (8), but unfortunately, they do not.[15] Hence, a statistical agency compiling a CPI has to choose between the two methods of index construction. The *Manual* advocated chaining as the "set of seasonal commodities which overlaps during two consecutive months is likely to be much larger than the set obtained by comparing the prices of any given month with a fixed base month" and "rapid sample attrition means that fixed base indices rapidly become unrepresentative and hence it seems preferable to use chained indices which can more closely follow marketplace developments." ILO (2004; 407). Thus the *Manual* recommended chained Fisher or Törnqvist indexes as target index concepts. As will be seen, this advice does not always work out too well.

If an index is to be produced at monthly or quarterly frequencies, it is not appropriate to use chaining when prices oscillate or "bounce".[16] This phenomenon can occur in the context of regular seasonal fluctuations or in the context of sales, resulting in *chain drift*; see the Appendix for a simple numerical example. Consider the following test due to Walsh (1901; 389), (1921; 540):[17]

$$Multiperiod \ Identity \ Test : P(p^1, p^2, q^1, q^2)P(p^2, p^3, q^2, q^3)P(p^3, p^1, q^3, q^1) = 1. \quad (9)$$

Thus price change is calculated over consecutive periods but an artificial final period is introduced where the prices and quantities revert back to the prices and quantities in the very first period. The test asks that the product of all of these price changes should equal unity. If prices have no definite trends but are simply bouncing up and down in a range,

---

[14]Fisher (1911; 203) introduced this fixed base and chain terminology. The concept of chaining is due to Lehr (1885) and Marshall (1887; 373).

[15]See Alterman, Diewert and Feenstra (1999; 61-65) for cases when the Törnqvist index P_T will satisfy the circularity test.

[16]See Szulc (1983)(1987), Hill (1988; 136-137) and Hill (1993; 388).

[17]This is Diewert's (1993; 40) term for the test. Walsh did not limit himself to just three periods as in (9); he considered an indefinite number of periods.

then (9) can be used to evaluate the amount of chain drift that occurs if chained indexes are used under these conditions. *Chain drift* occurs when an index does not return to unity when prices in the current period return to their levels in the base period; see the ILO (2004; 445).

There are at least three possible solutions to the chain drift that is associated with the use of a superlative index in a situation where monthly scanner data are available to the statistical agency for components of the CPI:[18] (i) Use an annual basket Lowe index that uses annual expenditure weights from a past year, which is the usual approach to CPI construction in most countries. (ii) Pick a base month and use fixed base superlative indexes relative to the chosen month. (iii) Use a Rolling Window multilateral index number approach adapted to the time series context.

The problem with the first method is that the Lowe index is subject to a small amount of upper level substitution bias, usually in the range of 0.15 to 0.40 percentage points per year.[19] The widespread use of the Lowe index is due to its practical nature and the fact that the amount of substitution bias is generally not all that large.[20] The problem with the second method is that picking a base month and calculating superlative indexes relative to this base is an asymmetric procedure and gives too much weight to the chosen base. Moreover, some commodities may not be available in the chosen base month and some commodities disappear in the later months of the comparison. Hence not all of the monthly price and quantity information is used in a fixed base method. Thus in the remainder of this paper, we will concentrate on the use of multilateral indexes as potential solutions to

---

[18]There is a possible fourth method to avoid chain drift within a year when using a superlative index and that is to simply compute a sequence of 12 year over year monthly indexes, so that say January prices in the previous year would be compared with January prices in the current year and so on. Handbury, Watanabe and Weinstein (2013) use this methodological approach for the construction of year over year monthly superlative Japanese consumer price indexes using the Nikkei point of sale data base. This data base has monthly price and expenditure data covering the years 1988 to 2010 and contains 4.82 billion price and quantity observations. This type of index number was recommended in the ILO (2004; chapter 22) as a valid year over year index that would avoid seasonality problems. However, central banks and other users require month to month CPIs in addition to year over year monthly CPIs and so the approach of Handbury, Watanabe and Weinstein does not solve the problems associated with the construction of superlative month to month indexes.

[19]For recent retrospective studies on upper level substitution bias for national CPIs, see Armknecht and Silver (2013), Diewert, Huwiler and Kohli (2009) and Huang, Wimalaratne and Pollard (2017). For studies of lower level substitution bias for a Lowe index, see Diewert, Finkel and Artsev (2009) and Diewert (2013a).

[20]Recent Canadian research has indicated that the substitution bias can be reduced substantially by more frequent updating of the annual basket; see Huang, Wimalaratne and Pollard (2017).

the chain drift problem.[21]

3 Multilateral Indexes

This section reviews some of the multilateral comparison methods that have been used in the scanner data context. These will be compared empirically in section 4 using artificial data sets where we know the truth by construction.

3.1 GEKS

The GEKS method for making international index number comparisons between countries is due to Gini (1931; 12), Eltetö and Köves (1964) and Szulc (1964).[22] Of course, it can also be adapted to make comparisons between multiple time periods.

The GEKS method in the time series context works as follows. Suppose we have price and quantity information for a component of the CPI on a monthly basis for a sequence of 13 consecutive months. Now pick one month (say month k) in this augmented year as the base month and construct Fisher price indexes for all 13 months relative to this base month. Denote the resulting sequence of Fisher indexes as $P_F(1/k), P_F(2/k), \ldots, P_F(13/k)$. The final set of GEKS indexes for the 13 months is simply the geometric mean of all 13 of the specific month indexes; i.e., the final set of *GEKS indexes for the months in the augmented year* is any normalization of the following sequence of indexes:[23]

$$\left[\prod_{k=1}^{13} P_F(1/k)\right]^{1/13}, \left[\prod_{k=1}^{13} P_F(2/k)\right]^{1/13}, \ldots, \left[\prod_{k=1}^{13} P_F(13/k)\right]^{1/13}. \tag{10}$$

The GEKS indexes in (10) have a number of important properties: (i) They satisfy Walsh's multiperiod identity test of equation (9) so that if any two months in the augmented year have exactly the same price and quantity vectors, then the index values will coincide

---

[21]For a recent study that takes a systematic look at the use of multilateral methods to create price indices from scanner data, see the Australian Bureau of Statistics (2016).

[22]It was derived in a different fashion by Eltetö and Köves (1964) and Szulc (1964) and thus the method is known as either the GEKS or EKS method for making multilateral comparisons.

[23]Balk (1981; 74) derived the GEKS parities using this type of argument rather than the usual least squares derivation of the GEKS parities; see Balk (1996) and Diewert (1999) for these alternative derivations.

for those two months; i.e., *the indexes are free from chain drift.* (ii) The indexes do not asymmetrically single out any single month to play the role of a base period; all possible base months contribute to the overall index values.[24] (iii) The indexes make use of all possible bilateral matches of the price data between any two months in the augmented year. That is, we are using what Ivancic, Diewert and Fox (2011; 26) called a "flexible basket" approach that incorporates new items as they become available. (iv) Strongly seasonal commodities make a contribution to the overall index values.

The last property explains why the augmented year should include at least 13 consecutive months, so that *strongly seasonal commodities* can make a contribution to the overall index.[25]

The major problem with the GEKS indexes defined by (10) is that they change as the data for a new month becomes available. A headline CPI cannot be revised from month to month due to the fact that many contracts are indexed to a country's headline consumer price index. A solution to this problem was proposed by Ivancic, Diewert and Fox (2011). Their method added the price and quantity data for the most recent month to the augmented year and dropped the oldest month from the old augmented year in order to obtain a new augmented year. The GEKS indexes for the new augmented year are calculated in the usual way and the ratio of the index value for the last month in the new augmented year to the index value for the previous month in the new augmented year is used as an *update factor* for the value of the index for the last month in the previous augmented year. The resulting indexes are called *Rolling Window GEKS indexes*, or for a thirteen month window, *Rolling Year GEKS indexes*.

Numerical experiments with Australian and Dutch scanner data from grocery chains showed that the Rolling Year GEKS indexes seem to work reasonably well; see Australian Bureau of Statistics (2016) Ivancic, Diewert and Fox (2011), de Haan and van der Grient (2011), and van der Grient and de Haan (2011). In particular, adding and dropping a month of data and recomputing the GEKS indexes does not seem to change past index

---

[24]Thus the GEKS procedure seems to be an improvement over the suggestion of Feenstra and Shapiro (2003) who chose only a single base month.

[25]A strongly seasonal commodity is one that is not present in the marketplace for all months of the year.

values very much.[26] Basically, the method seems to control chain drift quite well.[27]

There are two additional issues with using Rolling Window GEKS in the scanner data context: (i) How long should the window length be? The window length should be at least 13 months (or 5 quarters if a quarterly CPI is compiled) in order to allow for the influence of strongly seasonal commodities but what are the advantages and drawbacks of increasing this "standard" window length?[28] (ii) When a new window is computed, how should the index results from the new window be linked to the previous index values? As noted, 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, a method which has come to be known as the *movement splice*. 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*. de Haan (2015a; 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*.

We will return to these questions after we have studied our other multilateral methods.

---

[26]Balk (1981; 77) also observed the same phenomenon as he computed his GEKS indexes using successively larger data sets. Diewert (2013a) also found that Rolling Year GEKS estimates were quite close to their GEKS counterparts for his small data set on Israeli seasonal commodities.

[27]Ivancic, Diewert and Fox (2011; 33, footnote 19): "While a RWGEKS index, such as the RYGEKS, will not satisfy transitivity in practice and hence will be potentially subject to chain drift, comparisons within each window are transitive. Using this approach, chain drift is therefore unlikely to be a significant problem in any context likely to be faced by a statistical agency. Also, alternative approaches to linking the indexes could be investigated, such as using different overlapping periods for doing the linking, taking the geometric mean of overlapping comparisons in multiple windows, and so forth. The most obvious approach is pursued in this paper and works well in our empirical applications. An investigation into alternative approaches is left for future research." We return to this issue explicitly in section 11. The Australian Bureau of Statistics (2017) implemented the version of the Rolling Window GEKS index recommended by the current paper, from the fourth quarter of 2017 for twenty-eight expenditure classes, accounting for approximately 17% of the CPI weight as of April 2017. Statistics Netherlands computed RYGEKS indexes for some components of its CPI on an experimental basis with good results but they did not implement the method officially; see de Haan and van der Grient (2011) and de Haan (2015a) (2015b). Statistics New Zealand have implemented a version of RYGEKS, with adjustments for quality change, for consumer electronics scanner data; see Krsinich (2015).

[28]Melser (2018) suggested a weighted GEKS approach to resolve the issue of the optimal window length.

3.2   CCDI

The bilateral index Törnqvist index in (5) can be used to form a system of "star" indexes where we choose each period in turn as the base, calculate a sequence of indexes for each base and then take the geometric mean of the resulting star sequences. Thus a *preliminary price level for period t* using this methodology is π^t defined as follows:

$$\pi^t \equiv \left[\prod_{\tau=1}^{T} P_T(t/\tau) \right]^{1/T} = \left[\prod_{\tau=1}^{T} \prod_{n=1}^{N} \left(\frac{p_n^t}{p_n^\tau} \right)^{0.5(s_n^t + s_n^\tau)} \right]^{1/T} ; \quad t = 1, \dots, T. \quad (11)$$

The normalized sequence of price levels is defined as π^t/π^1 for $t = 1, \dots, T$. Obviously, this methodology is entirely analogous to Gini's methodology which led to the sequence of GEKS (unnormalized) price levels defined by (10). The only difference is that the Törnqvist bilateral price index formula, $P_T(t/\tau)$, is used as the basic building block rather than the Fisher formula, $P_F(t/\tau)$.[29]

The expressions (11) that define the (unnormalized) price levels π_t can be simplified. Define the sample average expenditure share for commodity n, $s_n^\bullet$, and the sample geometric mean of the prices for commodity n, $p_n^\bullet$, as follows:

$$s_n^\bullet \equiv \frac{1}{T} \sum_{t=1}^{T} s_n^t; \quad p_n^\bullet \equiv \prod_{t=1}^{T} \left(p_n^t \right)^{1/T}; \quad n = 1, \dots, N. \quad (12)$$

Definitions (12) can be used in order to define a new system of (unnormalized) price levels ρ^t for the T periods; i.e., define the logarithm of ρ^t as follows:

$$\rho^t \equiv \prod_{n=1}^{N} \left(\frac{p_n^t}{p_n^\bullet} \right)^{0.5(s_n^t + s_n^\bullet)} ; \quad t = 1, \dots, T. \quad (13)$$

The sequence of normalized or final price levels for the T observations in the sample is ρ^t/ρ^1 for $t = 1, \dots, T$. Note the similarity in structure of (5) and (13); the regular Törnqvist index $P_T(t/\tau)$ compares the level of prices in period t with the corresponding level in period whereas ρ^t can be thought of comparing the level of prices in period t with the corresponding

---

[29] de Haan and van der Grient (2011; 41) call the indexes (11) that use (5) as the bilateral building block GEKS-Törnqvist indexes. Fox and Syed (2016; 401) call the indexes defined by (11) CCD indexes. Caves, Christensen and Diewert (1982) used the GEKS methodology in the quantity context; i.e., they used bilateral Törnqvist quantity indexes as their basic building block rather than bilateral Törnqvist price indexes.

level of prices in an "average" observation which has expenditure shares equal to the sample average expenditure shares $s_n^\bullet$ and sample average prices $p_n^\bullet = \left[\prod_{t=1}^{T} p_n^t\right]^{1/T}$.

The price levels π^t defined by (11) are closely related to the price levels ρ^t defined by (13); i.e., substituting definitions (12) into (13) and using definitions (11) leads to the following relationship between the logarithms of π^t and ρ^t:

$$\ln \pi^t = \ln \rho^t + \alpha; \quad t = 1, \ldots, T \tag{14}$$

where the constant α is defined as follows:

$$\alpha \equiv \sum_{n=1}^{N} (1/2) s_n^\bullet \ln p_m^\bullet - (1/2T) \sum_{t=1}^{T} \sum_{n=1}^{N} s_n^t \ln p_n^t. \tag{15}$$

Equations (14) mean that the normalized price levels using the π^t and ρ^t coincide; i.e., we have:

$$\pi^t/\pi^1 = \rho^t/\rho^1; \quad t = 1, \ldots, T. \tag{16}$$

Comparing (11) to (13), it can be seen that making multilateral comparisons of prices across the T periods using the price levels ρ^t defined by (13) is analytically simpler than using the price levels π^t defined by (11). Caves, Christensen and Diewert (1982; 78) introduced the artificial country comparison idea into multilateral index number theory except that they introduced the idea in the context of making *quantity* comparisons across production units using distance functions. Inklaar and Diewert (2016; 429) extended the CCD methodology to making *price* comparisons across production units and derived equations (14). Thus we will call the multilateral price comparison method that uses the price levels ρ^t defined by (13) the *CCDI method.*

Using definitions (13) to define the (unnormalized) price levels ρ^t, it can be seen that the CCDI normalized price level in period t relative to period r, $(\rho^t/\rho^1)/(\rho^r/\rho^1)$, is independent of the choice of the normalizing period; i.e., $(\rho^t/\rho^1)/(\rho^r/\rho^1) = (\rho^t/\rho^s)/(\rho^r/\rho^s)$ for any period s.[30]

---

[30]The price levels ρ^t defined by (13) also satisfy the following version of Walsh's *multiperiod identity test* (9): for periods r, s and t, we have $(\rho^r/\rho^s)(\rho^t/\rho^r)(\rho^s/\rho^t) = 1$. The price levels ρ^t also satisfy the time reversal test and the circularity test. A less obvious test that the price level ρ^t also satisfies is the following price proportionality test: $\rho^t(\lambda p^t) = \lambda \rho^t(p^t)$ for arbitrary scalar $\lambda > 0$ where $\rho^t(p^t)$ is the function

The direct bilateral comparison of the price level of period t relative to period τ using the Törnqvist formula (5) depends only on the prices and expenditure shares pertaining to periods τ and t. The corresponding multilateral comparison of the price level in period t to period τ (in logarithms) is:[31]

$$\ln(\rho^t/\rho^\tau) = \sum_{n=1}^{N}(1/2)(s_n^t + s_n^\bullet)(\ln p_n^t - \ln p_n^\bullet) - \sum_{n=1}^{N}(1/2)(s_n^\tau + s_n^\bullet)(\ln p_n^\tau - ln p_n^\bullet) \quad (17)$$

Using (5), the prices of period t are compared directly to the prices in the base period whereas in (17), the prices in period t are compared to the prices in the artificial "average" period and then the prices in the "average" period are compared to the prices in the base period τ. Thus if there are smooth trends in prices and expenditure shares, chaining the bilateral indexes defined by (5) would probably be preferable (since chaining tends to reduce the spread between superlative indexes with smooth trends in the underlying price and quantity data)[32] whereas using the multilateral indexes defined by (17) essentially compares the prices in any two periods through the fixed base observation that has the sample average log prices and sample average expenditure shares. Of course, with monthly or weekly data, trends in prices and quantities are often far from smooth and under these nonsmooth conditions, the chain drift problem can become severe (as we shall see later) and the use of multilateral indexes is recommended.

3.3   WTPD

Ivancic, Diewert and Fox (2009) suggested adapting the weighted country product dummy multilateral method to the time series context for use with scanner data; this is now generally known as the Weighted Time Product Dummy (WTPD) method. Suppose that prices vary in an *approximately proportional manner* from period to period. Thus we have the following model:

$$p_n^t = a^t b_n e_n^t; \quad t = 1, \ldots, T; n = 1, \ldots, N. \quad (18)$$

$\rho^t(p^1, \ldots, p^T, q^1, \ldots, q^T)$ defined by (13) regarded as a function of the period t price vector, p^t. Note that the period t share vector s^t and the sample average share vector $s^\bullet \equiv [s_1^\bullet, \ldots, s_N^\bullet]$ also depends on p^t.

[31]Equation (17) can be used to establish the following bilateral identity test: if $p^t = p^\tau$ and $q^t = q^\tau$ (which implies $s^t = s^\tau$), then $\rho^t = \rho^\tau$.

[32]See Diewert (1978) on this point.

The parameter a^t can be interpreted as the price level for period t, b_n can be interpreted as a commodity n quality adjustment factor and e_n^t is a stochastic error term with mean 1. Define the logarithms of p_n^t and e_n^t as $y_n^t \equiv \ln p_n^t$ and $\varepsilon_n^t \equiv \ln e_n^t$ for $t = 1, \ldots, T; n = 1, \ldots, N$, define the logarithm of a^t as $\alpha^t \equiv \ln a^t$ for $t = 1, \ldots, T$ and define the logarithm of b_n as $\beta_n \equiv \ln b_n$ for $n = 1, \ldots, N$. Then taking logarithms of both sides of (18) leads to the following linear regression model:

$$y_n^t = \alpha^t + \beta_n + \varepsilon_n^t; \quad t = 1, \ldots, T; n = 1, \ldots, N. \tag{19}$$

The α^t and β_n can be estimated by solving a least squares minimization problem.[33] This is Summer's (1973) country product dummy multilateral method but with time periods in place of countries. Rao (2005) suggested a weighted-by-economic-importance version of Summer's method.[34] Applying weighted least squares means finding the α_t and β_n which solve the following *weighted least squares minimization problem*:

$$\min_{\alpha_1, \ldots, \alpha_T, \beta_1, \ldots, \beta_N} \sum_{t=1}^{T} \sum_{n=1}^{N} s_n^t (y_n^t - \alpha_t - \beta_n)^2. \tag{20}$$

Denote the β solution vector as β^*. The corresponding parameters b_n^* which appear in equations (18) are defined as the exponentials of the β_n^*; i.e., we have:

$$b_n^* \equiv \exp(\beta_n^*); \quad n = 1, \ldots, N. \tag{21}$$

Taking good N as a numeraire commodity, the quality adjustment parameter b_n^* essentially converts one unit of commodity n into the utility equivalent of b_n^* units of commodity N for $n = 1, 2 \ldots, N - 1$.[35]

---

[33] A normalization on the parameters such as $\alpha^1 = 0$ (which corresponds to $a^1 = 1$) is required to identify the parameters.

[34] See also Diewert (2004) (2005) (2012).

[35] This interpretation follows that of Rao (2005; 574) and de Haan (2015a; 21) (2015b; 6) but can be traced back to Summers (1973; 11).

The logarithm of the period t price level, α^{t*}, is then as follows:[36]

$$\alpha^{t*} \equiv \sum_{n=1}^{N} s_n^t (y_n^t - \beta_n^*). \tag{22}$$

Recall that $y_n^t = \ln p_n^t$ and $\beta_n^* = \ln(b_n^*)$ for each t and n. The period t (unnormalized) price level a_{t*} is the exponential of α^{t*}. Thus exponentiating both sides of (22) leads to the following formula for the (unnormalized) *weighted time product dummy price level* for period t:

$$a^{t*} = \prod_{n=1}^{N} \left(\frac{p_n^t}{b_n^*} \right)^{s_n^t}; \quad t = 1, \ldots, T. \tag{23}$$

Thus the unnormalized price level for period t is a period t expenditure share weighted geometric mean of the N period t quality adjusted prices p_n^t / b_n^*.

Define the *period t normalized weighted time product dummy price level π^t* by dividing a^{t*} by a^{1*}:

$$\pi^t \equiv a^{t*} / a^{1^*}; \quad t = 1, \ldots, T. \tag{24}$$

The Appendix considers some special cases for this index, which show that the WTPD multilateral method is recommended from the viewpoint of the economic approach to index number theory if (i) purchaser preferences are well approximated by Cobb-Douglas preferences, or (ii) purchaser preferences are well approximated by linear or additive preferences.[37]

Of particular note is a proof in the Appendix that the WTPD multilateral method is an *approximately additive method*.[38] The problem with an additive multilateral method is that, in general, it is not consistent with the economic approach to index number theory once the number of periods is greater than two; see Marris (1984; 52) and Diewert (1999; 49).

---

[36]Rao (1995)(2005; 577). See the Appendix for a fairly simple definition of the optimal β_n.

[37]In the context of the Constant Elasticity of Substitution (CES) model to be discussed later, the WTPD model is consistent with CES preferences if the elasticity of substitution is either 1 or plus infinity.

[38]We believe that this is a new result.

3.4   GK

The GK multilateral method was introduced by Geary (1958), and 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 the T (unnormalized) *price levels* $P^1, P^2, \ldots, P^T$ and a set of N *quality adjustment factors* $b_1, b_2, \ldots, b_N$. Let p^t and q^t denote the N dimensional price and quantity vectors for period t (with components p_n^t and q_n^t as usual) and let s^t (with components s_n^t) denote the period t expenditure share vector. Define the total consumption vector q over the entire window as the following simple sum of the period by period consumption vectors:

$$q \equiv \sum_{t=1}^{T} q^t \tag{25}$$

where $q \equiv [q_1, q_2, \ldots, q_N]$. The equations which determine the price levels P^t and quality adjustment factors b_n (up to a scalar multiple) are the following:

$$b_n = \sum_{t=1}^{T} \left[\frac{q_n^t}{q_n} \right] \left[\frac{p_n^t}{P^t} \right] ; \quad n = 1, \ldots, N; \tag{26}$$

$$P^t = p^t \cdot q^t / b \cdot q^t; \quad t = 1, \ldots, T \tag{27}$$

where $b \equiv [b_1, \ldots, b_N]$ is the vector of GK quality adjustment factors. The share of period t's purchases of commodity n in total sales of commodity n over all T periods can be defined as $S_n^t \equiv q_n^t / q_n$ for $n = 1, \ldots, N$ and $t = 1, \ldots, T$. Thus $b_n \equiv \sum_{t=1}^{T} S_n^t [p_n^t / P^t]$ is a (real) share weighted average of the inflation adjusted prices p_n^t / P^t for commodity n over all T periods. The period t quality adjusted sum of quantities sold, $b \cdot q^t = \sum_{n=1}^{N} b_n q_n^t$, is divided into the value of period t sales, $p^t \cdot q^t = \sum_{n=1}^{N} p_n^t q_n^t$, in order to obtain the period t (unnormalized) price level, P^t.

It can be seen that if a solution to equations (26) and (27) exists, then if all of the period price levels P^t are multiplied by a positive scalar λ say and all of the quality adjustment factors b_n are divided by the same λ, then another solution to (26) and (27)

is obtained. Hence, the b_n and P^t are only determined up to a scalar multiple and an additional normalization, such as $P^1 = 1$ or $b_N = 1$, is required to determine a unique solution to the system of equations defined by (26) and (27). It can also be shown that only $N + T - 1$ of the $N + T$ equations in (26) and (27) are independent.

Once the price levels P^t have been determined, the *real expenditure or volume* for period t, Q_t, can be defined as the period t value of sales, $p^t \cdot q^t$, divided by its period t price level, P^t:[39]

$$Q_t = p^t \cdot q^t / P^t = b \cdot q^t \quad \text{using (27)}; \quad t = 1, \dots, T \tag{28}$$

The second set of equations in (28) are the equations which characterize an *additive method*; i.e., aggregate quantity or volume in each period is a quality adjusted sum of the individual quantities. Hence an implication of the GK model is that it is consistent with utility maximizing purchasers who have a linear utility function. Thus the GK estimates can approximate the results that are based on a superlative index if substitutability between the products in the aggregate is very high.[40] In general, we would expect the GK price level estimates to suffer from at least some substitution bias. Furthermore, in general, the longer is the window length T, the bigger will be the substitution bias for at least some observations.

The usual method for obtaining a solution to equations (26) and (27) is to iterate between these equations. Thus set $b = 1_N$, a vector of ones and use equations (27) to obtain an initial sequence for the P^t. Substitute these P^t estimates into equations (26) and obtain a sequence of b_n estimates. Substitute these b_n estimates into equations (27) and obtain a new sequence of P^t estimates. Continue iterating between the two systems until convergence is achieved. In the Appendix, we derive an alternative method which is more efficient.

In summary: from the viewpoint of the economic approach to index number theory, we cannot endorse the use of the GK multilateral index in the time series context because it is exactly consistent with the economic approach only if purchaser preferences are either

---

[39]Khamis (1972; 101) also derived equations (28) in the time series context.

[40]We shall see that the GK indexes can also approximate the results that are based on a superlative index if substitutability between products is very low.

linear, so that all commodities in the class of commodities under consideration are perfect substitutes, or if purchaser preferences are Leontief, where there is a complete lack of substitutability between the commodities. The GK system is dominated by the WTPD multilateral system because the latter system is exact not only for linear preferences but also for Cobb-Douglas preferences.

3.5   Rolling Window Methods and the Linking Problem

Suppose that the CCDI multilateral index of (17) is used to construct a sequence of price levels, $\pi^1, \pi^2, \ldots, \pi^T$ for T consecutive periods where $T \geq 3$. The period t price level is defined as

$$\pi^t \equiv \rho^t / \rho^1; \quad t = 1, \ldots, T \tag{29}$$

where ρ^t is defined by (13). Data on period $T + 1$ expenditure shares s_n^{T+1} and prices p_n^{T+1} become available at the end of period $T + 1$ for $n = 1, \ldots, N$. A new set of multilateral indexes can now be constructed for the window of observations that include the data for periods $2, 3, \ldots, T, T + 1$. Define the average expenditure share for commodity n, $s_n^{\bullet *}$ and the average log price $p_n^{\bullet *}$ for commodity n as follows:

$$s_n^{\bullet *} \equiv \frac{1}{T} \sum_{t=2}^{T+1} s_n^t; \quad p_n^{\bullet *} \equiv \prod_{t=2}^{T+1} \left(p_n^t \right)^{1/T}; \quad n = 1, \ldots, N. \tag{30}$$

Definitions (30) can be used in order to define a new system of (unnormalized) CCDI price levels ρ^{t*} for the T periods in the new window; i.e., define ρ^{t*} as follows:

$$\rho^{t*} \equiv \prod_{n=1}^{N} \left(\frac{p_n^t}{p_n^{\bullet *}} \right)^{0.5(s_n^t + s_n^{\bullet *})}; \quad t = 2, \ldots, T, T + 1. \tag{31}$$

The new price levels need to be linked to the price levels generated by the original window of observations; i.e., we need to define a new price level for period $T + 1$ that extends the initial T price levels π^t defined by (29) for $t = 1, \ldots, T$. Suppose we link the new window price levels to the initial price levels at observation t where t could be any observation between periods 2 and T. Let $\pi^{T+1}(t)$ denote the resulting linked price level

for period $T + 1$ that depends on choosing period t as link period:

$$\pi^{T+1}(t) \equiv \left(\frac{\rho^t}{\rho^1} \right) \left(\frac{\rho^{T+1*}}{\rho^{t*}} \right); \quad t = 2, 3, \ldots, T \tag{32}$$

where the ρ^t are defined by (12) and (13) and the ρ^{t*} are defined by (30) and (31). Thus there are $T - 1$ possible choices of period t which can be used in (32) to link the price level for period $T + 1$ to the sequence of price levels $\pi^1, \ldots, \pi^T$ that were generated by the results of the initial window of observations; i.e., see definitions (29). Using the GEKS multilateral method rather than the CCDI method, Ivancic, Diewert and Fox (2009) (2011) suggested using period T as the linking observation. In the context of a somewhat different multilateral model, Krsinich (2016; 383) called this the *movement splice* method for linking the two windows. Krsinich (2013) (2016; 383) also suggested that a better choice of the linking observation in the context of her multilateral model was $t = 2$ and she called this the *window splice* method. de Haan (2015a; 26) suggested that the link period t should be chosen to be in the middle of the first window time span; i.e., choose $t = T/2$ if T is an even integer or $t = (T + 1)/2$ if T is an odd integer. The Australian Bureau of Statistics (2016; 12) called this the *half splice* method for linking the results of the two windows. However, without putting more structure on the underlying price and quantity data, it would seem that each choice of a linking period t running from $t = 2$ to $t = T$ is an equally valid choice of a period to link the two sets of price levels. Thus it would seem that we have $T - 1$ equally valid estimators for the period $T + 1$ price level, $\pi^{T+1}(t)$ defined by (32) for $t = 2, 3, \ldots, T + 1$, and so perhaps the "best" estimator for π^{T+1} is the *mean splice*, defined as the geometric mean of the $\pi^{T+1}(t)$:[41]

$$\pi^{T+1}(Mean) \equiv \left[\prod_{t=2}^{T} \pi^{T+1}(t) \right]^{1/(T-1)}. \tag{33}$$

In the context of very variable price data with little or no trends, choosing the mean estimator for π^{T+1} is probably a satisfactory strategy. However, we suggest an alternative strategy which may be superior.

---

[41]This method for linking the two windows was suggested by Ivancic, Diewert and Fox (2011; 33). Melser (2018) suggests an alternative method for linking which similarly treats each period symmetrically.

Suppose the price and quantity data for period $T + 1$ are identical to the price and quantity data for some period t such that $2 \leq t \leq T$; i.e., we have t such that:

$$s_n^{T+1} = s_n^t; \quad p_n^{T+1} = p_n^t; \quad n = 1, \dots, N. \tag{34}$$

Suppose further that we link the two windows at this observation t. The identity test for multilateral indexes requires that $\pi^{T+1}(t) = \pi^t$ where $\pi^{T+1}(t)$ is defined by (32) and π^t is defined by (29). Then, adapting formula (17) for the new window and using assumptions (34), their ratio can be written as follows:

$$\frac{\pi^{T+1}(t)}{\pi^t} = \frac{\prod_{n=1}^{N} \left(p_n^{T+1}/p_n^{\bullet*}\right)^{0.5(s_n^{T+1}+s_n^{\bullet*})}}{\prod_{n=1}^{N} \left(p_n^t/p_n^{\bullet*}\right)^{0.5(s_n^t+s_n^{\bullet*})}} = 1 \tag{35}$$

Thus if the price and quantity data for period $T + 1$ are exactly equal to the price and quantity data for period t, then linking the windows at observation t will preserve the identity test over the two windows; i.e., under these conditions it makes sense to link the two windows at period t. This logic carries over to situations where the price and quantity data of period t are closest (in some metric) to the price and quantity data of period $T + 1$; it makes sense to link the windows at the observation which has the most "similar" price and quantity data. If the data are reasonably smooth, then it is likely that the most similar price and quantity data will occur at observation T and thus under these conditions, movement splicing is justified.

The exact result (35) can be extended to the case where the price data of period $T+1$ is *proportional* to the price data of period t. Suppose the price and quantity data for period $T + 1$ are related to the data of period t (where $2 \leq t \leq T$) in the following way:

$$s_n^{T+1} = s_n^t; \quad p_n^{T+1} = \lambda p_n^t; \quad n = 1, \dots, N \tag{36}$$

where λ is a positive scalar. Suppose further that we link the two windows at this observation t. A *proportionality test* for multilateral indexes requires that $\pi^{T+1}(t) = \lambda \pi^t$ where $\pi^{T+1}(t)$ is defined by (32) and π^t is defined by (29). It can be shown that this property also holds when the CCDI multilateral method is used for each consecutive window and

the linking of the two windows is done at period t, where the data for periods t and $T+1$ satisfy (36).

These results suggest that linking of the two windows in a rolling window multilateral method be done at a period where the prices of period t are proportional (or close to being proportional) to the prices of period $T+1$. In order to implement this *fifth method* of linking the rolling windows, it is necessary to have measures of price proportionality between the two periods. Diewert (2009) suggested the following two measures of *relative price dissimilarity* between the prices of periods t and $T+1$:[42]

$$\triangle_{LQ}(p^t, p^{T+1}, q^t, q^{T+1}) \equiv \sum_{n=1}^{N}(1/2)(s_n^{T+1} + s_n^t)\left[\ln(p_n^{T+1}/P(p^t, p^{T+1}, q^t, q^{T+1})p_n^t)\right]^2;$$
(37)

$$\triangle_{AL}(p^t, p^{T+1}, q^t, q^{T+1}) \equiv \sum_{n=1}^{N}(1/2)(s_n^{T+1} + s_n^t)\{(p_n^{T+1}/P(p^t, p^{T+1}, q^t, q^{T+1})p_n^t)$$
$$+ (P(p^t, p^{T+1}, q^t, q^{T+1})p_n^t/p_n^{T+1}) - 2\}$$
(38)

where $P(p^t, p^{T+1}, q^t, q^{T+1})$ is a bilateral superlative index number formula.[43] $\triangle_{LQ}$ is the *weighted log quadratic index of relative price dissimilarity* and $\triangle_{AL}$ is the *weighted asymptotically linear index of relative price dissimilarity*. Both of these measures of relative price dissimilarity will equal their lower bound of 0 if prices in periods t and $T+1$ are proportional; i.e., if $p^{T+1} = \lambda p^t$ for some $\lambda > 0$. The bigger are the measures $\triangle_{LQ}$ and $\triangle_{AL}$, the less proportional are the prices in the two periods. In section 4.4, we will explain exactly how these measures of relative price dissimilarity can be used to link the windows using some test data.

4 Simulations with CES Preferences

The Australian Bureau of Statistics (2016) and the Dutch Central Bureau of Statistics[44] have done extensive computations using scanner data in order to compare alternative meth-

---

[42]See Fox, Hill and Diewert (2004) for the use of dissimilarity indexes in the context of detecting data outliers.

[43]Later in the paper, we will choose P to be P_T, the Törnqvist formula, when we use (37) and when we use (38), we will choose P to be P_F, the Fisher ideal price index.

[44]See de Haan and van der Grient (2011) and de Haan (2015a).

ods for aggregating prices at the first stage of aggregation. A problem with these comparisons is that it is not known which method is closest to the "truth". We will attempt to address this problem, at least in part, by constructing an artificial data set that is exactly consistent with purchasers having Constant Elasticity of Substitution (CES) preferences over a group of related items for a number of time periods. We will then determine how well the various multilateral methods studied earlier in this paper approximate the "true" indexes.

The CES functional form is another index number formula that is used in the first stage of price aggregation.[45] Using the economic approach to index number theory, it is assumed that purchaser preferences can be represented by a utility function $f(q)$ or by the unit cost function $c(p)$ which corresponds to the utility function. The *CES unit cost function* has the following functional form:

$$
\begin{aligned}
c(p_1, \ldots, p_N) &\equiv \left[\sum_{n=1}^{N} \alpha_n p_n^{1-\sigma} \right]^{1/(1-\sigma)} &&\text{if } \sigma \geq 0 \text{ and } \sigma \neq 1 \\
&\equiv \prod_{n=1}^{N} p_n^{\alpha_n} &&\text{if } \sigma = 1
\end{aligned}
\tag{39}
$$

where σ and the α_n are positive parameters, with $\sum_{n=1}^{N} \alpha_n = 1$. The unit cost function defined by (39) corresponds to *a Constant Elasticity of Substitution (CES) utility function* which was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961). The parameter σ is the *elasticity of substitution*, which is a constant for all pairs of commodities; when $\sigma = 0$, the unit cost function defined by (39) becomes linear in prices and hence corresponds to a fixed coefficients or Leontief utility function which exhibits 0 substitutability between all commodities. When $\sigma = 1$, the corresponding aggregator or utility function is a Cobb-Douglas function. When σ approaches $+\infty$, the corresponding aggregator function f approaches a linear aggregator function which exhibits infinite substitutability between each pair of inputs. The CES unit cost function defined by (39) is *not* a fully flexible functional form (unless the number of commodities N being aggregated is 2)

---

[45] From January 2015, the Bureau of Labor Statistics uses this functional form to construct elementary indexes in the initial estimates of the Chained Consumer Price Index for All Urban Consumers; see Bureau of Labor Statistics (2018).

but it is frequently used to aggregate commodities in a group of goods which are thought to be highly substitutable with each other.

Suppose that we are given the price vectors $p^t \equiv [p_1^t, \ldots, p_N^t]$ for $t = 1, .., T$. Then if purchasers have CES preferences and are minimizing the costs of achieving their utility levels in each period, it will turn out that the components of their period t expenditure share vectors $s^t \equiv [s_1^t, \ldots, s_N^t]$ for $t = 1, .., T$ will be equal to the following expressions:

$$s_n^t = \frac{\alpha_n (p_n^t)^{1-\sigma}}{\sum_{n=1}^{N} \alpha_n (p_n^t)^{1-\sigma}}; \quad n = 1, \ldots, N; t = 1, \ldots, T. \tag{40}$$

Thus given the price vectors p^t, the vector of positive parameters $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$ whose components sum to unity and the nonnegative parameter σ where $\sigma \neq 1$, then the share vectors s^t can be computed using equations (40) for $t = 1, \ldots, T$. If in addition we are given total expenditures on the N commodities for period t, say $e^t > 0$ for $t = 1, \ldots, T$, then we can compute the components of the CES quantity vector for period t, say $q^t = [q_{t1}, \ldots, q_{tN}]$ for $t = 1, \ldots, T$ using the following equations:

$$q_n^t \equiv e^t s_n^t / p_n^t; \quad n = 1, \ldots, N; t = 1, \ldots, T. \tag{41}$$

Finally, the CES (unnormalized) *period t price levels*, P^t, are computed as follows:

$$P^t \equiv c(p^t) = \left[\sum_{n=1}^{N} \alpha_n (p_n^t)^{1-\sigma} \right]^{1/(1-\sigma)}; \quad t = 1, \ldots, T. \tag{42}$$

The corresponding CES (unnormalized) *period t quantity levels* are defined as follows:

$$Q_t \equiv e^t / P^t = p^t \cdot q^t / P^t; \quad t = 1, \ldots, T. \tag{43}$$

Of course, the CES normalized price levels can be defined as $\pi^t \equiv P^t / P^1$ for $t = 1, \ldots, T$. We will use this method for constructing an artificial data set that is consistent with CES preferences in the following section.

4.1   Artificial Data From CES Preferences

We now assume that $T = 12$ and $N = 4$, and that the parameters α_n which appear in (39) are defined as follows: $\alpha \equiv [\alpha_1, \alpha_2, \alpha_3, \alpha_4] \equiv [0.2, 0.2, 0.2, 0.4]$. The elasticity of substitution parameter σ will take on the values 0, 0.5, 1, 2, 4, 10 and 20. In the scanner data context, it is likely that it is between 1 and 5.[46]

In order to calculate the period t (unnormalized) CES price level $P^t \equiv c(p^t)$, where $c(p)$ is defined by (39), we require information on the period t vector of prices, $p^t \equiv [p_{t1}, p_{t2}, p_{t3}, p_{t4}]$ for $t = 1, \ldots, 12$. These prices are listed in Table 1. Once the unnormalized CES price levels P^t have been constructed, the normalized CES price levels can be calculated as $\pi^t_{CES} \equiv P^t/P^1$ for $t = 1, \ldots, 12$. In order to calculate the CES quantity vectors that match up with the given CES price vectors, we also require information on total expenditures e^t on the four commodities for each period t. These exogenously given expenditures are also listed in Table 1.

Table 1: Price and Expenditure Data for the Artificial Data Set

t	p_{t1}	p_{t2}	p_{t3}	p_{t4}	e^t
1	2.00	1.00	1.00	0.50	10
2	1.75	**0.50**	0.95	0.55	13
3	1.60	1.05	0.90	0.60	11
4	1.50	1.10	0.85	0.65	12
5	1.45	1.12	**0.40**	0.70	15
6	1.40	1.15	0.80	0.75	13
7	1.35	1.18	0.75	0.70	14
8	1.30	**0.60**	0.72	0.65	17
9	1.25	1.20	0.70	0.70	15
10	1.20	1.25	**0.40**	0.75	18
11	1.15	1.28	0.70	0.75	16
12	1.10	1.30	0.65	0.80	17

Note: Sales prices are in bold type.

The prices of commodities 1 and 3 trend downward throughout the sample period while the prices of commodities 2 and 4 trend upward. The trends in commodities 1 and 4 are very smooth but the trends in commodities 2 and 3 are interrupted by sales: item 2 goes on sale in periods 2 and 8 and item 3 goes on sale in periods 5 and 10. Total expenditures e^t on the commodity group trend upwards except in the four periods after a sale when

---

[46]See the empirical evidence on the magnitude of σ using Australian scanner data in Ivancic, Diewert and Fox (2010).

aggregate expenditures fall a bit.

For each value of the elasticity of substitution σ, we can evaluate the period t expenditure share vectors $s^t \equiv [s_1^t, s_2^t, s_3^t, s_4^t]$ using the p^t vectors that are listed in Table 1 and equations (40). Once the share vectors have been constructed, the components of the corresponding quantity vectors $q^t \equiv [q_{t1}, q_{t2}, q_{t3}, q_{t4}]$ can be constructed using equations (41).

Once the set of twelve price vectors p^t and twelve quantity vectors q^t have been constructed for the alternative values, we can compare the period t chained Fisher, chained Törnqvist, Weighted Time Product Dummy, Geary-Khamis, GEKS and CCDI normalized price levels ($\pi_{FCH}^t, \pi_{TCH}^t, \pi_{WTPD}^t, \pi_{GK}^t, \pi_{GEKS}^t, \pi_{CCDI}^t$ respectively) to the corresponding true CES normalized price level, π_{CES}^t. These alternative indexes are evaluated for σ equal to 0, 0.5, 1, 2, 4, 10 and 20 and are listed in Table A5 of the Appendix, and plotted in Figure 1 for the considered values of σ up to 10.[47]

Figure 1: Alternative Price Levels for Different Methods and Elasticities of Substitution
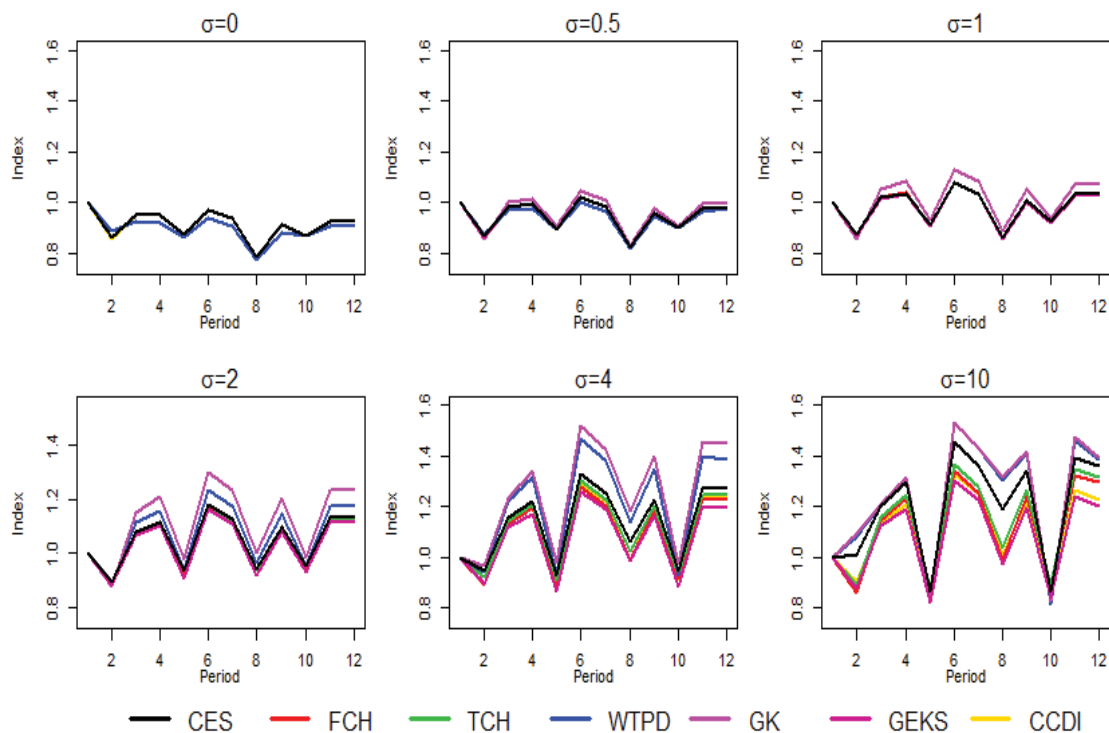
When $\sigma = 0$, the CES, chained Fisher (FCH), GK and GEKS indexes all coincide. Furthermore the chained Törnqvist (TCH) and CCDI indexes are so close to the CES that they cannot be distinguished from each other in Figure 1. The only outliers in this case

---

[47]To simplify programming, $\sigma = 1.001$ was actually used throughout for the "$\sigma = 1$" case.

are the WTPD indexes which lie below the other indexes after period 2. Preferences which allow for no substitution between products are not realistic, but the results provide a check on our computations.

For $\sigma = 0.5$, the GK indexes have a modest upwards bias while the WTPD indexes have a modest downward bias as compared to the true cost of living index. The remaining indexes are all very close to the CES price levels and cannot be distinguished separately in Figure 1.

When $\sigma = 1$ the chained Törnqvist, CCDI and WTPD price levels should be exactly equal to the corresponding CES price levels and this expectation is borne out. The chained Fisher and GEKS indexes are also very close to the corresponding CES indexes and cannot be distinguished from each other in Figure 1. The GK price levels turn out to be substantially above the other indexes after period 2.

When $\sigma = 2$, the GK price levels are substantially above the corresponding WTPD levels after period 2 and the WTPD price levels are well above the CES price levels. Thus these indexes have substantial amounts of substitution bias for our numerical example. The chained Törnqvist and CCDI price levels are very close to the corresponding CES price levels while the chained Fisher and GEKS price levels end up being slightly below the CES price levels.

When $\sigma = 4$, the GK and WTPD price levels are substantially above the CES price levels for most observations.[48] Thus these indexes have substantial amounts of substitution bias for our numerical example in this case. The chained Törnqvist and CCDI price levels are 2-3% under the corresponding CES price levels at the end of the sample period while the chained Fisher and GEKS price levels end up being 4-5% below the CES price levels at the end of the sample period. Thus none of our approximations to the true CES indexes end up being very close at the end of the sample period but the upward bias in the GK and WTPD indexes is very large indeed (13.8% and 8.8% respectively).

When $\sigma = 10$, the WTPD and GK price levels are virtually identical[49] and these

---

[48] The vector of WTPD quality adjustment factors is $b_{WTPD} \equiv [1.84, 1.26, 0.93, 1.00]$ and the vector of GK quality adjustment factors is $b_{GK} \equiv [1.78, 1.16, 0.87, 1.00]$.

[49] The new vector of WTPD quality adjustment factors is $b_{WTPD} \equiv [1.66, 0.93, 0.98, 1.00]$ and the vector of GK quality adjustment factors is $b_{GK} \equiv [1.64, 0.92, 0.97, 1.00]$. Thus the two additive methods are converging to the same preferences as σ becomes large.

indexes turn out to give the best approximation to the true CES indexes although for many observations, they lie well above the corresponding true indexes. The remaining indexes lie below the CES indexes, with the chained Törnqvist and chained Fisher indexes coming the closest to the corresponding CES indexes.

From Table A5 of the Appendix, when $\sigma = 20$, the WTPD and GK price levels are again virtually identical.[50] For most observations, these price levels are either very close to the corresponding CES price levels or are above them, although at the end of the sample they are slightly below. The remaining indexes are generally substantially below the CES price levels. Thus for very high elasticities of substitution, the WTPD and GK methods provide the best approximation to CES preferences, as might be expected due to their being additive methods; see sections 3.3 and sec:GK.

In summary: for elasticities of substitution in the most likely range of 1 to 4, the four methods based on the use of bilateral superlative indexes approximate CES preferences reasonably well with the chained Törnqvist generally doing the best. The GK indexes have substantial upward biases in all cases while the WTPD indexes also have substantial upward biases when equals 2 or 4, but they are unbiased when $\sigma = 1$.

4.2 Artificial Data Adjusted for Sales

The CES indexes which were constructed in section 4.1 did not suffer from a chain drift problem; i.e., if prices and quantities in any two periods were exactly the same, the CES index levels for those two periods would also be the same. However, in actual practice, the assumptions which the Konüs (1924) true cost of living methodology relies on are not met: when a commodity goes on sale at say one half of its regular price, consumers tend to purchase amounts that not only satisfy their needs for the sale period but also they stock up to partially satisfy their needs for the subsequent period. This violates a basic assumption that is used in the economic approach to index number theory; i.e., the economic approach to index number theory assumes that purchases of goods made during a period coincide with the consumption of these purchased goods within the period. As was explained earlier,

---

[50]The new vector of WTPD quality adjustment factors is $b_{WTPD} \equiv [1.64, 0.91, 0.99, 1.00]$ and the vector of GK quality adjustment factors is $b_{GK} \equiv [1.64, 0.91, 0.99, 1.00]$. Thus the two additive methods have converged to the same preferences when $\sigma = 20$.

this stockpiling activity will cause chained superlative indexes to drift downwards.[51]

To model this situation, we will use the price and expenditure data listed in Table 1 and the assumption of CES preferences for varying σ as in section 4.1 but the quantity data for commodities that go on sale will be adjusted. Thus commodity 2 goes on sale in periods 2 and 8. For periods 3 and 9, we take the predicted quantities for these items that are generated by the CES model described in section 4.1 and we somewhat arbitrarily adjust these quantities downward to half of the predicted levels that were generated by the CES model. Similarly, commodity 3 goes on sale in periods 5 and 10. Again, for periods 6 and 11, we adjust the quantities of these commodities to half of the predicted levels that were generated by the CES model in section 4.1 Thus for periods $t = 3, 6, 9$ and 11 and for each alternative value of σ, we have a new q^t vector and so for these periods, we need to define a new level of expenditure $e^t \equiv p^t \cdot q^t$ and new expenditure shares $s_n^t \equiv p_n^t q_n^t / e^t$ for $t = 3, 6, 9$ and 11 and $n = 1, 2, 3, 4$.

As an illustration, for $\sigma = 2$ the resulting prices, quantities and expenditures for commodity 2 are plotted in figure 2, where it can be seen that the sales periods generate spikes in quantities sold and in the corresponding expenditures.

Thus for the eight periods $1, 2, 4, 5, 7, 8, 10$ and 12, the unnormalized and normalized CES price levels, P^t and $\pi_{CES}^t \equiv P^t / P^1$ for the new data set, are the same as in section 4.1. For periods 3, 6, 9 and 11, there are no CES price levels but for convenience, in Table A6 and Figure 3, we simply use the normalized CES price levels π_{CES}^t from section 4.1 for all periods.

Once the new set of twelve price vectors p^t and twelve quantity vectors q^t have been constructed for the alternative σ values,[52] we can compare the period t chained Fisher, chained Törnqvist, fixed base Fisher, fixed base Törnqvist, Weighted Time Product Dummy, Geary-Khamis, GEKS and CCDI normalized price levels (π_{FCH}^t, π_{TCH}^t, π_{FFB}^t, π_{TFB}^t, π_{WTPD}^t, π_{GK}^t, π_{GEKS}^t, and π_{CCDI}^t, respectively) to the corresponding true CES normalized price level,

---

[51]Szulc (1983) (1987) was perhaps the first modern price statistician to note the chain drift problem in the context of sales. However, Persons (1921; 110) using his data set on crop production noted the downward drift of Fisher chained indexes relative to a Fisher index that measured price change directly between periods 1 and T and he also applied Walsh's (1901; 389) multiperiod identity test to demonstrate this downward drift. Frisch (1936; 9) noted Persons' research in his survey paper on index numbers.

[52]The new p^t and q^t will coincide with the p^t and q^t that were used in section 4.1 except for periods 3, 6, 9 and 11, we use new q^t.

Figure 2: Sales Adjusted Data, Commodity 2
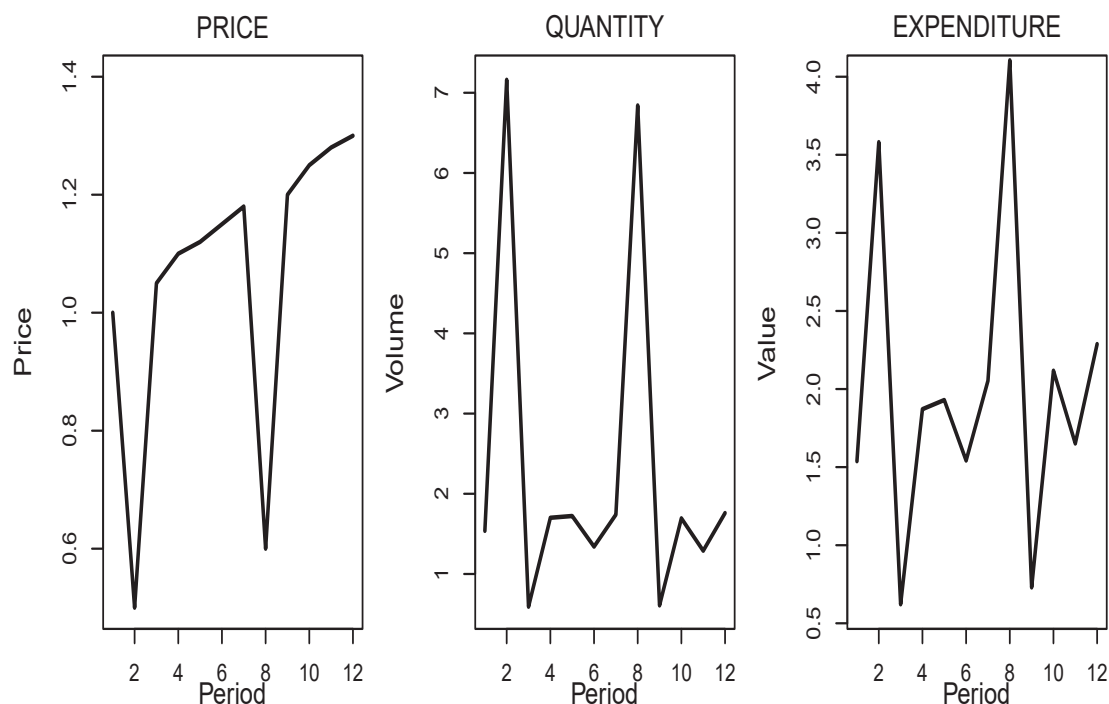
π_{CES}^t, for all periods except periods 3, 6, 9 and 11. These alternative indexes are evaluated for σ equal to 0, 0.5, 1, 2, 4, and 10 and are listed in Table A6 of the Appendix, and plotted in Figure 3.

Using this adjusted data set, the chained Fisher and chained Törnqvist indexes have substantial downward biases as compared to the corresponding true CES indexes (excluding periods 1, 3, 6, 9 and 11). The new data set illustrates that superlative indexes are indeed subject to substantial downward chain drift when products are sold at substantially discounted prices.

From Figure 3 (and Table A6 of the Appendix), for $\sigma = 0$, it can be seen that the fixed base Fisher (FFB) and GK price levels coincide with the CES price levels for all periods except periods 3, 6, 9 and 11 where the CES indexes are not well defined for the new data.[53] The fixed base Törnqvist (not plotted in Figure 3), GEKS and CCDI price levels are very close to the corresponding CES price levels. The WTPD price levels have a small amount of downward bias while the chained Fisher and the chained Törnqvist (not plotted in Figure 3) price levels have substantial downward biases.

---

[53]Recall that we are using the CES normalized price levels from section 4.1 for all periods in the present section in order to simplify the charts for this section.

Figure 3: Alternative Price Levels for Sales Adjusted Data
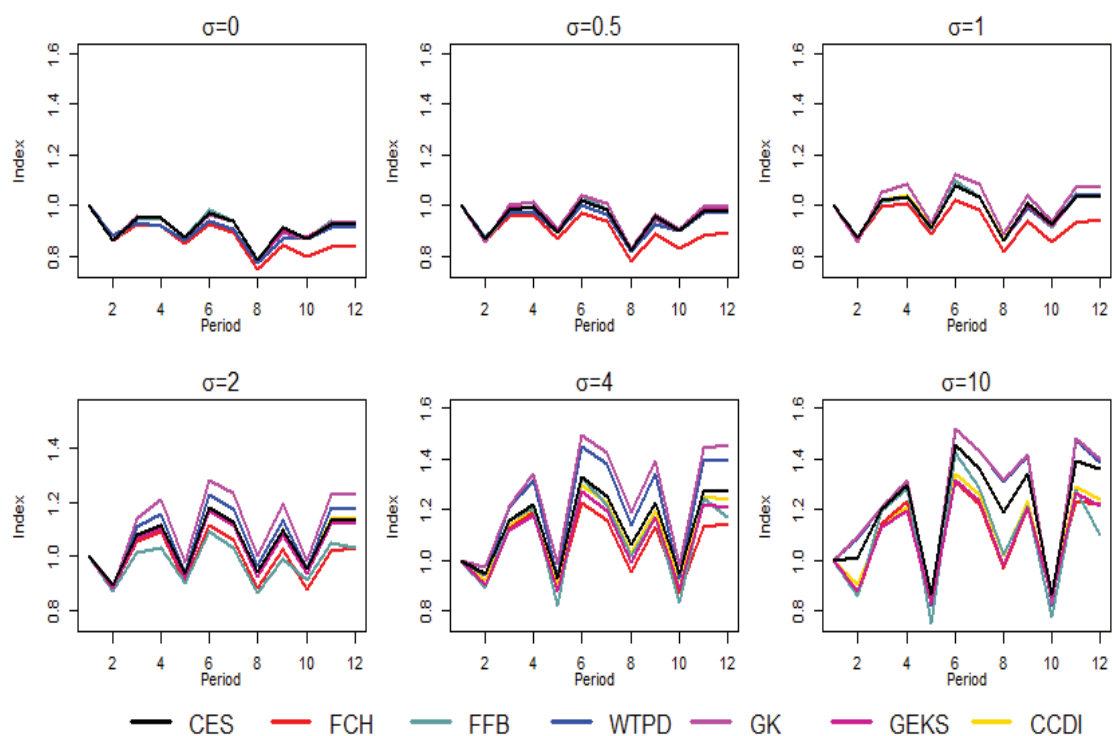
When $\sigma = 0.5$, the fixed base Fisher (and fixed base Törnqvist), GEKS and CCDI price levels are all very close to the corresponding CES price levels. The GK indexes are slightly above and the WTPD indexes are slightly below the corresponding CES index levels. The two chained superlative indexes end up well below the other indexes; i.e., the chain drift problem for superlative indexes is severe for this artificial data set. We move on to the more applicable σ equals 1, 2 and 4 results.

When $\sigma = 1$, the two chained superlative indexes have a large downward bias compared to the target CES index. The GK price levels are substantially above the corresponding CES price levels. The remaining indexes are all fairly close to their CES counterparts. The fixed base Törnqvist indexes are exactly equal to their CES counterparts (except for observations 3, 6, 9 and 11 when the CES indexes are not applicable); see Table A6.

When $\sigma = 2$, the fixed base Törnqvist and CCDI price levels are very close to their CES counterpart price levels. The GEKS price level for period 12 ends up 1.5% below the corresponding CES price level while the fixed base Fisher price level ends up 2.5% below the target index. The chained Fisher and chained Törnqvist indexes end up 9.0% and 11.5% below the target index in period 12 while the WTPD and GK price levels for period

12 end up 3.2% and 8.5% above the CES price level.

When $\sigma = 4$, the CCDI indexes are closest to their CES counterparts and the CCDI price level for period 12 ends up 2.7% below the corresponding CES level. The fixed base Törnqvist and GEKS price levels end up 4.4% and 5.5% below the CES price level at period 12. The fixed base Fisher, the chained Törnqvist and the chained Fisher indexes all have substantial downward biases and end up 8.3%, 10.2% and 10.3% below the corresponding CES price level for period 12. The WTPD and GK indexes have substantial upward biases and end up 9.0% and 14.2% above the corresponding CES price level at period 12.

When $\sigma = 10$, all of the alternative indexes have some significant biases for some periods. The WTPD and GK price levels are close to each other and are substantially above their CES counterparts for many observations. These price levels end up being only 1.7% and 2.5% above the corresponding CES price level for period 12. The worst performing indexes are the fixed base Fisher and fixed base Törnqvist indexes which end up 19.4% and 16.8% below the CES price level for period 12. The chained Fisher and Törnqvist indexes end up 10.2% and 9.4% below the CES price level for period 12. Finally the GEKS and CCDI price levels end up 11.0% and 9.1% below the CES price level for period 12. However, it should be kept in mind that in most empirical applications, the elasticity of substitution will be well below 10.

Some tentative conclusions that can be drawn from the material presented in this section are as follows: (i) Chained superlative indexes are not useful target indexes for a CPI when dealing with aggregating scanner data where discounted prices are prevalent. (ii) The CCDI multilateral method worked best overall for our numerical example for elasticities of substitution in the range $0 \leq \sigma \leq 4$. (iii) GK indexes had substantial upward biases relative to the corresponding CES true cost of living price levels for elasticities of substitution in the range $1 \leq \sigma \leq 4$. (iv) Weighted Time Product Dummy indexes will work well if $\sigma = 1$ or if $\sigma \geq 10$ but for our example, they had substantial upward biases for elasticities of substitution in the range $2 \leq \sigma \leq 4$.

We turn now to an examination of the problems associated with linking the results from one window of observations to another subsequent window.

4.3 Single Window versus Linked CCDI Price Levels

In this section, using the sales adjusted data, we draw on the results in section 3.5 and compare the single window π_{CCDI}^{t} price levels which were listed in section 4.2 for various data sets that depended on the elasticity of substitution with the corresponding CCDI price levels that use linked windows. The single window results were based on a window that consisted of all 12 observations. These CCDI price levels of course depended on σ and we denote their period 12 values as $\pi_{CCDI}^{12}(\sigma)$ in this section. For each of our σ values, we will form two windows of length 11. The first such window will consist of the data for periods $1 - 11$ and the second window will consist of the data for periods $2 - 12$. In keeping with CPI methodology which does not allow revisions to occur for the flagship CPI, we assume that the final linked series for periods 1 to 11 will consist of the CCDI normalized price levels using the first window of length 11. The linked in CPI for period 12 will be formed by using equations (32) in section 3.5 to link in the results from the second window by choosing the link periods $t = 2, 3, \ldots, 11$. Denote these alternative linked normalized price levels for period 12 by $\pi_{L}^{12}(t, \sigma)$ for choice of link period $t = 2, 3, \ldots, 11$. Define the geometric mean of these ten alternative period 12 price levels as $\pi_{L}^{12}(Mean, \sigma)$.

We first compare the period 12 linked normalized price levels $\pi_{L}^{12}(t, \sigma)$ to the period 12 single window normalized price level, $\sigma_{CCDI}^{12}(\sigma)$. Define the differences between these period 12 price levels as follows:

$$D(t, \sigma) \equiv \pi_{L}^{12}(t, \sigma) - \pi_{CCDI}^{12}(\sigma); \quad t = 2, 3, \ldots, 11; \sigma = 0, 0.5, 1, 2, 4, 10. \quad (44)$$

We also define the difference between the period 12 single window CCDI price level for each σ, $\pi_{CCDI}^{12}(\sigma)$, and the mean period 12 price level using the geometric mean of the 10 alternative link periods, $\pi_{L}^{12}(Mean, \sigma)$, as follows:

$$D(Mean, \sigma) \equiv \pi_{L}^{12}(Mean, \sigma) - \pi_{CCDI}^{12}(\sigma); \quad \sigma = 0, 0.5, 1, 2, 4, 10. \quad (45)$$

If these differences are large in magnitude, then this indicates a chain drift problem with

the use of successive CCDI linked windows.[54] Table 2 lists these differences.

For σ in the range $0 \leq \sigma \leq 2$, it can be seen that the differences in the period 12 CCDI price levels due to the choice of alternative link periods are small. Choosing the link period to be period 4 or 10 leads to virtually zero differences but the differences are also small when periods 2, 5, 7 or 8 are chosen. Choosing the mean of the ten period specific estimates also leads to a fairly low difference between the single window period 12 CCDI price level and the two window linked period 12 price levels.

The differences are much larger when the true cost of living index elasticity of substitution is equal to 4 or 10. It can be seen that the differences are smallest when the linking period is chosen to be periods 2, 5 or 10. Note that sales occurred during each of these periods. When $\sigma = 10$, the period 12 linked CCDI price level is always at least 1 percentage point above its single window CCDI price level for period 12 and for the choice of period 6 as the link observation, it is 3 percentage points above.

Table 2: Differences at Period 12, $D(t, \sigma)$, between the Single Window CCDI Price Levels and the Linked CCDI Price Levels as Functions of the Linking Period t and the Elasticity of Substitution

t	$D(t,0)$	$D(t,0.5)$	$D(t,1)$	$D(t,2)$	$D(t,4)$	$D(t,10)$
2	0.00030	0.00014	0.00021	0.00067	0.00212	0.01004
3	-0.00197	-0.00149	-0.00098	0.00050	0.00567	0.02197
4	-0.00001	0.00011	0.00021	0.00098	0.00603	0.02640
5	-0.00029	0.00000	0.00021	0.00006	-0.00222	0.01370
6	0.00154	0.00206	0.00265	0.00442	0.01035	0.03114
7	0.00002	0.00011	0.00021	0.00092	0.00537	0.02639
8	0.00041	0.00022	0.00021	0.00125	0.00645	0.02581
9	-0.00177	-0.00137	-0.00098	0.00014	0.00451	0.02312
10	-0.00015	0.00003	0.00021	0.00010	-0.00202	0.01647
11	0.00151	0.00204	0.00265	0.00432	0.00946	0.02675
Mean	-0.00004	0.00019	0.00046	0.00133	0.00457	0.02216

We now compare the period 12 linked normalized price levels $\pi_L^{12}(t, \sigma)$ to the period 12 true cost of living index, $\pi_{CES}^{12}(\sigma)$. Define the *biases* between these period 12 price levels as follows:

$$B(t, \sigma) \equiv \pi_L^{12}(t, \sigma) - \pi_{CES}^{12}(\sigma); \quad t = 2, 3, \ldots, 11; \sigma = 0, 0.5, 1, 2, 4, 10. \quad (46)$$

---

[54]If the Törnqvist bilateral indexes satisfied the circularity test exactly, then these differences would all be 0 and there would be no chain drift problem with the use of chained Törnqvist indexes. However, as was seen in section 4.2, the circularity test does not hold exactly and there is a chain drift problem.

We also define the difference between the period 12 CES price level for each σ, $\pi_{CES}^{12}(\sigma)$, and the mean period 12 price level using the geometric mean of the ten alternative link periods, $\pi_L^{12}(Mean, \sigma)$, as follows:

$$B(Mean, \sigma) \equiv \pi_L^{12}(Mean, \sigma) - \pi_{CES}^{12}(\sigma); \quad \sigma = 0, 0.5, 1, 2, 4, 10. \qquad (47)$$

For each σ, these differences measure the bias in the various two window CCDI period 12 price levels compared to the corresponding period 12 true cost of living indexes, $\pi_{CES}^{12}(\sigma)$. Table 3 lists these differences. It can be seen that the bias in the two window CCDI price levels at period 12 for elasticities of substitution in the range $0 \leq \sigma \leq 4$ fairly small for all choices of the linking point but smallest if periods 2, 4, 5, 7, 8 or 10 are chosen. Note that the bias is largest in magnitude if periods 3, 6, 9 or 11 are chosen as the linking observations. These four periods are the periods that immediately follow a sale. Since in real life, a sale could occur during any period, this result indicates that it will be difficult to come up with a general rule as to which period in the first window is the "best" period to choose in terms of minimizing bias. Under these circumstances, choosing the mean splice method may be a conservative strategy which would avoid large biases.

Table 3: Biases at Period 12, $B(t, \sigma)$, as Functions of the Linking Period t and the Elasticity of Substitution

t	$B(t, 0)$	$B(t, 0.5)$	$B(t, 1)$	$B(t, 2)$	$B(t, 4)$	$B(t, 10)$
2	-0.00002	0.00068	0.00249	0.00187	-0.03189	-0.11420
3	-0.00230	-0.00094	0.00130	0.00169	-0.02833	-0.10227
4	-0.00034	0.00065	0.00249	0.00218	-0.02797	-0.09784
5	-0.00061	0.00054	0.00249	0.00126	-0.03622	-0.11054
6	0.00122	0.00260	0.00494	0.00562	-0.02365	-0.09310
7	-0.00031	0.00065	0.00249	0.00212	-0.02864	-0.09785
8	0.00008	0.00076	0.00249	0.00244	-0.02756	-0.09843
9	-0.00210	-0.00083	0.00130	0.00134	-0.02949	-0.10112
10	-0.00048	0.00057	0.00249	0.00130	-0.03603	-0.10777
11	0.00118	0.00258	0.00494	0.00551	-0.02455	-0.09749
Mean	-0.00037	0.00073	0.00274	0.00253	-0.02944	-0.10208

Table 3 shows that for large elasticities of substitution, the bias pattern has shifted: linking at observations 6 and 11 now reduces the bias. When $\sigma = 10$, the bias is greatest in magnitude when the linking observation is chosen to be observations 2, 5 or 10. Note that observations 2, 5 and 10 are periods when sales occurred.

Our conclusion at this point is that no clear pattern emerged as to which observation in the window was the "best" one for linking the second window to the first. Until more conclusive evidence on this problem is obtained, it seems advisable to use the mean splice.

4.4   Relative Price Similarity Linking

As was suggested in section 3.5, it is likely that price comparisons between two periods will be more accurate if the structure of relative prices is more similar.[55] 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 (2009), Sergeev (2009), Aten and Heston (2009) and Diewert (2009).[56]

We use the two measures of relative price dissimilarity defined by (37) and (38), $\triangle_{AL}$ (the weighted asymptotic linear index of relative price dissimilarity), and $\triangle_{LQ}$ (the weighted log quadratic index).[57] These measures are nonnegative and the bigger they are, the more dissimilar are the relative prices for periods r and t. Note that the measures defined by (37) and (38) require that a bilateral index formula, $P(p^r, p^t, q^r, q^t)$, be specified so that these measures can be calculated. For the asymptotic linear measure of relative price dissimilarity, we will use the Fisher price index P_F defined by (4) and for the log quadratic measure, we will use the Törnqvist index P_T defined by (5) as the bilateral indexes that appear in (37) and (38) respectively. We use the same price and quantity data that were used in sections 4.1 and 4.2; see the Appendix for details of the application.

In Table 4, we list the mean absolute difference between our ten approximating indexes to the corresponding true cost of living indexes, $\pi_{CES}^t(\sigma)$ for each of our representative values. However, we exclude periods 3, 6, 9 and 11 from this comparison because the true cost of living is not defined for these observations.

---

[55]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 e.g. Hill (1999) (2009) and Diewert (2013b). Hill (2001) (2004) also pursued this similarity of relative prices approach in the time series context.

[56]See Diewert (2009) for a discussion of the relative merits of the various measures.

[57]This measure is a weighted generalization of the nonproportionality measure suggested by Allen and Diewert (1981; 433).

Table 4: Mean Absolute Differences in Percentage Points between $\pi_{CES}^t(\sigma)$ and Ten Approximating Indexes as Functions of the Elasticity of Substitution

σ	B_{FCH}	B_{TCH}	B_{FFB}	B_{TFB}	B_{WTPD}	B_{GK}	B_{GEKS}	B_{CCDI}	B_{AL}	B_{LQ}
0	3.78	4.77	0.00	0.11	1.66	0.00	0.12	0.08	0.55	0.63
0.5	3.89	4.84	0.12	0.04	0.89	1.19	0.05	0.17	0.47	0.43
1	4.06	4.81	0.46	0.00	0.00	2.58	0.17	0.29	0.53	0.27
2	4.81	4.75	1.91	0.37	2.19	5.68	1.40	0.10	1.07	0.47
4	6.84	5.65	5.96	3.41	5.37	8.90	4.98	2.47	2.69	1.61
10	9.08	7.68	10.83	9.19	4.91	5.01	9.57	7.83	6.05	5.08

When $\sigma = 0$, the fixed base Fisher and GK indexes are exactly equal to the CES indexes. The fixed base Törnqvist, CCDI and GEKS indexes are all very close to the true cost of living index. When $\sigma = 0.5$, GEKS does best followed closely by the fixed base Fisher and CCDI indexes. When $\sigma = 1$, the fixed base Törnqvist and WTPD price levels are exactly equal to the corresponding CES price levels. The GEKS, CCDI and LQ price similarity linked price levels are all fairly close to the corresponding CES price levels. For $\sigma = 2$, The CCDI price levels are very close to the corresponding CES price levels followed by the fixed base Törnqvist and the log quadratic price similarity linked indexes (LQ indexes). The GK price levels are on average furthest from the corresponding CES price levels when $\sigma = 2$ and 4. For $\sigma = 4$, the LQ indexes are closest to the corresponding true cost of living indexes followed by the CCDI and Asymptotic Linear price similarity linked indexes (AL indexes). Finally, for $\sigma = 10$, the WTPD, GK and LQ indexes are the clear winners.

The results can be summed up as follows:

- The Chained Fisher and Chained Törnqvist indexes performed poorly for all elasticities of substitution.

- The Weighted Time Product Dummy indexes worked well for our numerical example when the elasticity of substitution was equal to 1 or 10 but they did not work well when was equal to 2 or 4.

- The Geary-Khamis indexes worked well when $\sigma = 0$ or 10 but poorly when $\sigma = 1, 2$ or 4.

- For $0 \leq \sigma \leq 0.5$, the Fixed Base Fisher, Fixed Base Törnqvist, GEKS and CCDI indexes all worked well. However the cases where 1 are the cases of interest.

- For $1 \leq \sigma \leq 2$, the CCDI indexes performed well.

- The LQ price similarity linked indexes performed the best by a substantial margin for $\sigma = 4$ and the LQ generally performed well for $1 \leq \sigma \leq 10$.

In applications of multilateral methods using scanner data at the elementary level of aggregation, it is likely that the elasticity of substitution is greater than one so the very limited results in this section support the use of the CCDI and Log Quadratic price similarity linked indexes in this context.[58]

We have not directly addressed the possibility of using the price dissimilarity measures for choosing the period for linking windows in applying multilateral methods. However, we can make some observations on this from the price dissimilarity measures of Appendix tables A3 and A4. Using the Log Quadratic measures from Table A4, $\triangle_{LQ}^{12t}$ suggests the optimality of using period 11 as the link period for updating an initial 11 period window as period 12 data becomes available, with the smallest dissimilarity value being 0.003. This suggests that the movement splice updating method is optimal, i.e. using the immediately prior period as the linking period. However, if we have, say, an initial 10 period window, as period 11 data becomes available, $\triangle_{LQ}^{11t}$ suggests the linking period should be period 9, which means that movement splice updating is not optimal.

What this example demonstrates is that it is not necessarily the case that, for any window length, the previous period is the optimal linking period. Again, in the absence of additional information, this suggests the use of the mean splice.

5 Conclusion

The dot points at the end of the previous section will not be repeated here but they could serve as the main conclusions.

There are two important limitations of our study. First, while we believe this paper

---

[58]We have not recommended the use of the Sato-Vartia formula to compute price indexes that are consistent with CES preferences because a limitation of the CES functional form is that the degree of substitutability between every pair of commodities is constrained to be the same. Methods like GEKS, CCDI or similarity linking are based on the use of superlative bilateral index number formulae as building blocks and thus these methods will allow for an arbitrary pattern of substitutability between each commodity pair.

has significantly advanced understanding, more research into how the different multilateral methods perform under different conditions is needed. Second, we have assumed that all prices and quantities are positive over all periods. Our results will still be valid if some quantities are zero in some periods but the assumption of positive prices cannot be entirely relaxed. Many of the multilateral methods depend on taking the logarithms of prices and if prices are 0 in some periods, this can cause the multilateral method to be undefined. The economic approach to index number theory can deal with new and disappearing commodities using the shadow price approach due to Hicks (1940) which can in theory generate positive reservation prices to go along with the associated zero quantities. However, it will be difficult for statistical agencies to estimate these reservation prices.

An important result is that linking the price and quantity data for a new period to the data of previous periods by using a price dissimilarity measure is the only multilateral method that is consistent with Walsh's powerful *multiperiod identity test*; i.e., if the price data for the new period is equal to or proportional to the price data of a previous period, then the price level for the new period will be equal to or proportional to the price level of the previous period. The drawback to the use of this multilateral method is that it requires agreement on how to measure the degree of price dissimilarity. For our sales adjusted data sets, we found that the Log Quadratic measure of price dissimilarity worked well. It remains to be seen whether this result will hold up under other conditions.

In the meantime, for elasticities of substitution in reasonable ranges that are expected to be found empirically, overall our results suggest the use of the CCDI index combined with a new method, the mean splice, for updating.

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 (2013), "Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indexes", *Review of Income and Wealth*, forthcoming.

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.

Australian Bureau of Statistics (2017), "An Implementation Plan to Maximise the Use of Transactions Data in the CPI", Information Paper 6401.0.60.004, 16 June, Canberra: ABS.

Australian Bureau of Statistics (2018), "6401.0 - Consumer Price Index, Australia, Dec 2017", Media Release, 31 January 2018, http://www.abs.gov.au/ausstats/abs@.nsf/lookup/6401.0Media%20Release1Dec%202017

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.

Baltas, G. (2001), "Utility-consistent Brand Demand Systems with Endogenous Category Consumption: Principles and Marketing Applications", *Decision Sciences* 32, 399-421.

Bureau of Labor Statistics (2007), *BLS Handbook of Methods: Ch 17. The Consumer Price Index.* Bureau of Labor Statistics, Washington DC. https://www.bls.gov/opub/hom/pdf/homch17.pdf

Bureau of Labor Statistics (2018), "Improving initial estimates of the Chained Consumer Price Index", February, Washington D.C. https://www.bls.gov/opub/mlr/2018/article/improving-initial-estimates-of-the-chained-consumer-price-index.htm

Brynjolfsson, E., W.E. Diewert, F. Eggers, K.J. Fox and A. Gannamaneni (2018), "The Digital Economy, GDP and Consumer Welfare: Theory and Evidence", paper presented at the ESCoE Conference on Economic Measurement, Bank of England, 16-17 May.

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.

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. (1993), "The Early History of Price Index Research", pp. 33-65 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1995), On the Stochastic Approach to Index Numbers", Discussion Paper 95-31, Department of Economics, University of British Columbia, Vancouver, Canada.

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.

Diewert, W.E. (2013a), "An Empirical Illustration of Index Construction using Israeli Data on Vegetables" paper presented at the 13th Meeting of the Ottawa Group On Prices, May 1-3, Copenhagen, Denmark.

Diewert, W.E. (2013b), "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", Vancouver School of Economics Discussion Paper 17-10, University of British Columbia.

Diewert, W.E., Y. Finkel and Y. Artsev (2009), "Empirical Evidence on the Treatment of Seasonal Products: The Israeli Experience", pp. 53-78 in *Price and Productivity*

Measurement: Volume 2: Seasonality , W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Victoria, Canada: Trafford Press.

Diewert, W.E., M. Huwiler and U. Kohli (2009), "Retrospective Price Indices and Substitution Bias", *Swiss Journal of Economics and Statistics* 145(2), 127-135.

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.

Fox, K.J., R.J., Hill and W.E. Diewert (2004), "Identifying Outliers in Multi-Output Models", *Journal of Productivity Analysis* 22, 73-94.

Fox, K.J. and I. Syed (2016), "Price Discounts and he Measurement of Inflation*", Journal of Econometrics* 191, 398-406.

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.

Goolsbee, A.D. and P.J. Klenow (2018), "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce", *American Economic Review, Papers and Proceedings* 108, 488-492.

van der Grient, H.A. and J. de Haan (2011), "Scanner Data Price Indexes: The 'Dutch Method" versus Rolling Year GEKS", paper presented at the 12th Meeting of the Ottawa Group, May 4-6, Wellington, New Zealand.

de Haan, J. (2008), "Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data", paper presented at the EMG Workshop 2008, University of New South Wales, Sydney, Australia, December 10-12.

de Haan, J. (2015a), "A Framework for Large Scale Use of Scanner Data in the Dutch CPI", Paper presented at the 14th meeting of the Ottawa Group, May 22, Tokyo. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf

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 H.A. van der Grient (2011), "Eliminating Chain drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161, 36-46.

Hanbury, J., T. Watanabe and D.E. Weinstein (2013), "How Much do Official Price Indexes Tell us about Inflation", paper presented at the Economic Measurement Group Workshop Asia 2013, Tokyo, Japan, October 16.

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

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

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

Hill, T.P. (1993), "Price and Volume Measures", pp. 379-406 in *System of National Accounts 1993*, Eurostat, IMF, OECD, UN and World Bank, Luxembourg, Washington, D.C., Paris, New York, and Washington, D.C.

Huang, N., W. Wimalaratne and B. Pollard (2017), "The Effects of the Frequency and Implementation Lag of Basket Updates on the Canadian CPI", *Journal of Official Statistics*, forthcoming.

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

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

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

Ivancic, L., W.E. Diewert and K.J. Fox (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.

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

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. (2013), "Using the Rolling Year Time Product Dummy Method for Quality Adjustment in the Case of Unobserved Characteristics", Ottawa Group 2013 Meeting, Copenhagen, Denmark, May 1-3.

Krsinich, F. (2015), "Implementation of Consumer Electronics Scanner Data in the New Zealand CPI", Statistics New Zealand. Paper presented at the New Zealand Association of Economists conference, Wellington, New Zealand, 3 July.

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.

Lehr, J. (1885), *Beitrage zur Statistik der Preise*, Frankfurt: J.D. Sauerlander.

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.

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

Melser, D. (2018), "Scanner Data Price Indexes: Addressing Some Unresolved Issues", *Journal of Business & Economic Statistics* 36, 516-522.

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.

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.

Sato, K. (1976), "The Ideal Log Change Index Number", *Review of Economics and Statistics* 78, 121-126.

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.

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 förhällandet mellan finska markens och svenska kronans köpkraft?", *Ekonomiska Samfundets Tidskrift* 39, 1-39 reprinted as pp. 121-160 in *Collected Scientific Papers of Leo Törnqvist*, Helsinki: The Research Institute of the Finnish Economy, 1981.

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

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.

Online Appendix

A Numerical Example of the Chain Drift Problem

Suppose that we are given the price and quantity data for two commodities for four periods. The data are listed in Table A1.

Table A1: Price and Quantity Data for Two Products for Four Periods

t	p_1^t	p_2^t	q_1^t	q_2^t
1	1.0	1.0	10	100
2	0.5	1.0	5000	100
3	1.0	1.0	1	100
4	1.0	1.0	10	100

The first commodity is subject to sales (in period 2), when the price drops to half of its normal level of 1. In period 1, we have a "normal" off-sale demand for commodity 1 which is equal to 10 units. In period 2, the sale takes place and demand explodes to 5000 units.[59] In period 3, the commodity is off sale and the price is back to 1 but most shoppers have stocked up in the previous period so demand falls to only 1 unit. Finally in period 4, the commodity is off sale but we are back to the "normal" demand of 10 units. Commodity 2 is dull: its price is 1 in all periods and the quantity sold is 100 units in each period. Note that the only thing that has happened going from period 3 to 4 is that the demand for commodity one has picked up from 1 unit to the "normal" level of 10 units. Also note that, conveniently, the period 4 data are exactly equal to the period 1 data so that for Walsh's test to be satisfied, the product of the period to period chain links must equal one.

Table A2 lists the fixed base Fisher, Laspeyres and Paasche price indexes, $P_{F(FB)}$, $P_{L(FB)}$ and $P_{P(FB)}$ and as expected, they behave perfectly in period 4, returning to the period 1 level of 1. Then the chained Fisher, Törnqvist, Laspeyres and Paasche price indexes, $P_{F(CH)}$, $P_{T(CH)}$, $P_{L(CH)}$ and $P_{P(CH)}$ are listed. Obviously, the chained Laspeyres and Paasche indexes have chain drift bias that is extraordinary but what is interesting is that the chained Fisher has a 2% downward bias and the chained Törnqvist has a close to 3% downward bias.[60]

If these data were monthly, and they repeated themselves three times over the year, the overall chain drift bias would build up to the 6 to 8% range, which is significant.

The problem is this: when commodity 1 comes off sale and goes back to its regular price in period 3, *the corresponding quantity does not return to the level it had in period 1*: the period 3 demand is only 1 unit whereas the "normal" period 1 demand for commodity 1

---

[59]This example is taken from Diewert (2012). It is based on a similar example due to de Haan and van der Grient (2011; 39), and draws on actual movements observed in Dutch scanner data. When the price of a detergent product went on sale at approximately one half of the regular price, the volume sold shot up approximately one thousand fold; see de Haan (2008). This paper brought attention to the magnitude of volume fluctuations due to sales.

[60]Feenstra and Shapiro (2003) also looked at the chain drift problem that was caused by sales and restocking dynamics. Their suggested solution was to use fixed base indexes.

Table A2: Fixed Base and Chained Fisher, Törnqvist, Laspeyres and Paasche Indexes

t	$P_{F(FB)}$	$P_{L(FB)}$	$P_{P(FB)}$	$P_{F(CH)}$	$P_{T(CH)}$	$P_{L(CH)}$	$P_{P(CH)}$
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.698	0.955	0.510	0.698	0.694	0.955	0.510
3	1.000	1.000	1.000	0.979	0.972	1.872	0.512
4	1.000	1.000	1.000	0.979	0.972	1.872	0.512

was 10 units. It is only in period 4 that demand for commodity one recovers to the period 1 level. However, since prices are the same in periods 3 and 4, all of the chain links show no change (even though quantities are changing) and this is what causes the difficulties. If demand for commodity one in period 3 had immediately recovered to its "normal" period 1 level of 10, then there would be no chain drift problem.

B Relationships Between Bilateral and Multilateral Indexes

The difference between the direct Törnqvist index (5) and the corresponding multilateral index (17) simplifies to the following expression:[61]

$$
\begin{aligned}
&\ln P_T(t/\tau) - \ln(\rho_t/\rho_\tau) \\
&= \sum_{n=1}^{N}(1/2)(s_{\tau n} - s_{\bullet n})(\ln p_{tn} - \ln p_{\bullet n}) - \sum_{n=1}^{N}(1/2)(s_{tn} - s_{\bullet n})(\ln p_{\tau n} - \ln p_{\bullet n}). \quad \text{(B1)}
\end{aligned}
$$

This formula can be used to show that if $T = 2$, then $\ln P_T(2/1) = \ln(\rho_2/\rho_1)$ so that the direct Törnqvist bilateral index $P_T(2/1)$ equals the multilateral CCDI index ρ_2/ρ_1 if there are only two periods in the multilateral comparison.

Comparing the WTPD price level for period t relative to period $\tau, \pi^t/\pi^\tau$, with the corresponding Törnqvist bilateral index, $P_T(t/\tau)$, defined by (5). Using (D18), we have:

$$
\ln(\pi^t/\pi^\tau) = \sum_{n=1}^{N} s_n^t(\ln p_n^t - \beta_n^*) - \sum_{n=1}^{N} s_n^\tau(\ln p_n^\tau - \beta_n^*). \quad \text{(B2)}
$$

Using (5) and (B2), it can be shown that:

$$
\begin{aligned}
\ln P_T(t/\tau) - \ln(\pi^t/\pi^\tau) &= \sum_{n=1}^{N}(1/2)(s_n^\tau - s_n^t)(y_n^t + y_n^\tau - 2\beta_n^*) \\
&= \sum_{n=1}^{N}(s_n^\tau - s_n^t)[(1/2)\ln(p_n^t/b_n^*) + (1/2)\ln(p_n/b_n^*)] \quad \text{(B3)}
\end{aligned}
$$

---

[61]Formula (B1) shows that if $p^t = p^\tau$ and $q^t = q^\tau$ (which implies $s^t = s^\tau$), then $P_T(t/\tau) = \rho_t/\rho_\tau = 1$. Formula (B1) also shows that if expenditure shares are constant over time so that $s_{tn} = s_{\bullet t}$ for all t and n, then $P_T(t/\tau) = \rho_t/\rho_\tau$.

where the b_n^* are defined by (21). It is difficult to determine the sign of the right hand side of (B3) except in the case (D23) where expenditure shares are constant across all observations in which case the right hand side of (B3) is equal to zero and thus $P_T(t/\tau)$ equals π^t/π^τ in this case.

Using (17) and (B2), it is possible to compare the logarithm of the CCDI price level for period t relative to period τ, $\ln(\rho^t/\rho^\tau)$, to the logarithm of the WTPD price level for period t relative to period τ, $\ln(\pi^t/\pi^\tau)$:

$$
\begin{aligned}
\ln(\rho^t/\rho^\tau) - \ln(\pi^t/\pi^\tau) &= \sum_{n=1}^{N}(1/2)(s_n^t + s_n^\bullet)(\ln p_n^t - \ln p_n^\bullet) \\
&- \sum_{n=1}^{N}(1/2)(s_n^\tau + s_n^\bullet)(\ln p_n^\tau - \ln p_n^\bullet) - \left[\sum_{n=1}^{N} s_n^t(\ln p_n^t - \beta_n^*) - \sum_{n=1}^{N} s_n^\tau(\ln p_n^\tau - \beta_n^*)\right] \\
&= \sum_{n=1}^{N}(s_n^\tau - s_n^\bullet)\left[(1/2)\ln(p_n^\tau/b_n^*) + (1/2)\ln(p_n^\bullet/b_n^*)\right] \\
&- \sum_{n=1}^{N}(s_{tn} - s_n^\bullet)\left[(1/2)\ln(p_n^t/b_n^*) + (1/2)\ln(p_n^\bullet/b_n^*)\right].
\end{aligned}
\tag{B4}
$$

Again, it is difficult to determine the sign of the right hand side of (B4) except in the case (D23) where expenditure shares are constant across all observations, in which case the right hand side of (B4) is equal to zero and thus ρ^t/ρ^τ equals π^t/π^τ in this case.

C The Weighted Time Product Dummy Method: Estimation Details

The first order necessary (and sufficient) conditions for solving (20) are the following T equations (C5) and the N equations (C6):

$$
\alpha_t + \sum_{n=1}^{N} s_n^t \beta_n = \sum_{n=1}^{N} s_n^t y_n^t; \quad t = 1, \ldots, T;
\tag{C5}
$$

$$
\sum_{t=1}^{T} s_n^t \alpha_t + \left(\sum_{t=1}^{T} s_n^t\right) \beta_n = \sum_{t=1}^{T} s_n^t y_n^t; \quad n = 1, \ldots, N.
\tag{C6}
$$

Multiply both sides of equation $t = 1$ in equations (C5) by s_{11}, equation $t = 2$ in (C5) by $s_{21}, \ldots$, equation $t = T$ in (C5) by s_{T1}, sum the resulting equations and subtract this summed equation from both sides of the first equation in (C6). The resulting equation has no α_t terms. Do similar summations of equations (C5) to eliminate the terms $\sum_{t=1}^{T} s_n^t \alpha_t$

in each of the equations (C6). The resulting system of N equations is the following one:

$$
\sum_{t=1}^{T} s_1^t \beta_1 - \sum_{t=1}^{T} (s_1^t)^2 \beta 1 - \sum_{t=1}^{T} s_1^t s_2^t \beta_2 - \sum_{t=1}^{T} s_1^t s_3^t \beta_3 - \ldots - \sum_{t=1}^{T} s_1^t s_n^t \beta_N
$$

$$
= \sum_{t=1}^{T} s_1^t y_1^t - \sum_{t=1}^{T} (s_1^t)^2 y_1^t - \sum_{t=1}^{T} s_1^t s_2^t y_2^t - \sum_{t=1}^{T} s_1^t s_3^t y_3^t - \ldots - \sum_{t=1}^{T} s_1^t s_n^t y_n^t;
$$

$$
\vdots
$$

$$
\sum_{t=1}^{T} s_n^t \beta_N - \sum_{t=1}^{T} s_n^t s_1^t \beta_1 - \sum_{t=1}^{T} s_n^t s_2^t \beta_2 - \sum_{t=1}^{T} s_n^t s_3^t \beta_3 - \ldots - \sum_{t=1}^{T} (s_n^t)^2 \beta_N
$$

$$
= \sum_{t=1}^{T} s_n^t y_n^t - \sum_{t=1}^{T} s_n^t s_1^t y_1^t - \sum_{t=1}^{T} s_n^t s_2^t y_2^t - \sum_{t=1}^{T} s_n^t s_3^t y_3^t - \ldots - \sum_{t=1}^{T} (s_n^t)^2 y_n^t.
$$

$$(C7)$$

Note that $s_1^t = 1 - \sum_{n=2}^{N} s_n^t$, $s_2^t = 1 - s_1^t - \sum_{n=3}^{N} s_n^t \ldots$, $s_n^t = 1 - \sum_{n=1}^{N-1} s_n^t$. Substituting these equations into the N equations (C7) leads to the following equivalent system of equations:

$$
\sum_{t=1}^{T} s_1^t (s_2^t + \ldots s_n^t) \beta_1 - \sum_{t=1}^{T} s_1^t s_2^t \beta_2 - \sum_{t=1}^{T} s_1^t s_3^t \beta_3 - \ldots - \sum_{t=1}^{T} s_1^t s_{N-1}^t \beta_{N-1} - \sum_{t=1}^{T} s_1^t s_n^t \beta_N
$$

$$
= \sum_{t=1}^{T} s_1^t s_2^t (y_{t1} - y_{t2}) + \sum_{t=1}^{T} s_1^t s_3^t (y_{t1} - y_{t3}) + \ldots + \sum_{t=1}^{T} s_1^t s_n^t (y_{t1} - y_n^t);
$$

$$
\vdots
$$

$$
\sum_{t=1}^{T} s_n^t s_1^t \beta_1 - \sum_{t=1}^{T} s_n^t s_2^t \beta_3 - \ldots - \sum_{t=1}^{T} s_n^t s_{N-1}^t \beta_{N-1} + \sum_{t=1}^{T} s_n^t (s_1^t + \ldots s_{N-1}^t) \beta_N
$$

$$
= \sum_{t=1}^{T} s_n^t s_1^t (y_n^t - y_{t1}) + \sum_{t=1}^{T} s_n^t s_2^t (y_n^t - y_{t2}) + \ldots + \sum_{t=1}^{T} s_n^t s_{N-1}^t (y_{tN} - y_{tN-1}). \qquad (C8)
$$

Recall that $y_n^t \equiv \ln p_n^t$. Thus the right hand side of the first equation in (C8) is a weighted sum of the logarithmic price ratios $\ln(p_{t1}/p_{t2}), \ln(p_{t1}/p_{t3}), \ldots, \ln(p_{t1}/p_n^t)$. Similarly, the right hand side of the last equation in (C8) is a weighted sum of the logarithmic price ratios $\ln(p_n^t/p_{t1}), \ln(p_n^t/p_{t2}), \ldots, \ln(p_n^t/p_{tN-1})$.

Equations (C8) can be simplified if we define the following TN^2 weights w_{tnj}:

$$
\begin{aligned}
w_{tnj} &\equiv s_n^t s_j^t > 0; \quad n \neq j; n = 1, \ldots, N; j = 1, \ldots, N; t = 1, \ldots, T; \\
w_{tnn} &\equiv 0; \quad n = 1, \ldots, N; t = 1, \ldots, T.
\end{aligned}
\qquad (C9)
$$

Define the N^2 weights w_{nj} as follows:[62]

$$
w_{nj} \equiv \sum_{t=1}^{T} w_{tnj}; \quad n = 1, \ldots, N; j = 1, \ldots, N. \qquad (C10)
$$

---

[62]Note that $w_{nn} = 0$ for $n = 1, \ldots, N$.

Using these definitions, equations (C8) simplify to the following N equations:

$$\left(\sum_{j=1}^{N} w_{nj}\right)\beta_n - \sum_{j=1}^{N} w_{nj}\beta_j = \sum_{t=1}^{T}\sum_{j=1}^{N} w_{tnj}y_n^t - \sum_{t=1}^{T}\sum_{j=1}^{N} w_{tnj}y_{tj}; \quad n = 1,\ldots,N. \qquad \text{(C11)}$$

Equations (C11) can be further simplified if we divide both sides of equation n in (C11) by $\sum_{j=1}^{N} w_{nj}$ for $n = 1,\ldots,N$. The resulting system of equations is:

$$\beta_n - \sum_{j=1}^{N} f_{nj}\beta_j = \sum_{t=1}^{T}\sum_{j=1}^{N} f_{tnj}(y_n^t - y_{tj}); \quad n = 1,\ldots,N \qquad \text{(C12)}$$

where the fractions $f_{nj} \geq 0$ and $f_{tnj} \geq 0$ are defined as follows:[63]

$$f_{nj} \equiv w_{nj}/\sum_{k=1}^{N} w_{nk}; \quad n, j = 1,\ldots,N; \qquad \text{(C13)}$$

$$f_{tnj} \equiv w_{tnj}/\sum_{k=1}^{N} w_{nk}; \quad n, j = 1,\ldots,N; t = 1,\ldots,T. \qquad \text{(C14)}$$

It can be verified that $\sum_{j=1}^{N} f_{nj} = 1$ for $n = 1,\ldots,N$ and $\sum_{t=1}^{T}\sum_{j=1}^{N} f_{tnj} = 1$ for $n = 1,\ldots,N$. Let $F \equiv [f_{nj}]$ be the N by N matrix which has element nj equal to f_{nj} for $n = 1,\ldots,N$ and $j = 1,\ldots,N$. Let $f \equiv [f_1,\ldots,f_N]^T$ where $f_n \equiv \sum_{t=1}^{T}\sum_{j=1}^{N} f_{tnj}(y_n^t - y_{tj})$ for $n = 1,\ldots,N$.[64] Then the N equations in (C12) can be written using matrix notation as:

$$[I_N - F]\beta = f \qquad \text{(C15)}$$

where I_N is the N by N identity matrix and $\beta \equiv [\beta_1,\ldots,\beta_N]^T$ is the column vector of the β_n. Thus it would appear that the solution to the weighted least squares minimization problem (20) is equal to $\beta^* \equiv [I_N - F]^{-1}f$. However, the inverse of the matrix $[I_N - F]$ does not exist since its columns are linearly dependent; i.e., we have $[I_N - F]1_N = 0_N$ where 1_N is a vector of ones and 0_N is a vector of zeros. The solution to this problem is simple: set one of the β_n equal to a definite number and drop one of the equations in the system of N equations defined by (C15).[65] In our empirical work, we will choose to set β_N equal to 0 and to drop the last equation in (C15). Under these conditions, the remaining betas can

---

[63]$f_{nn} = 0$ for $n = 1,..,N$ and $f_{tnn} = 0$ for $n = 1,\ldots,N$ and $t = 1,\ldots,T$. The remaining fractions are positive under our assumptions that each price p_n^t and quantity q_n^t is positive for all n and t.

[64]Note that the elements of the matrix F depend on just expenditure shares whereas the elements of the vector f are share weighted averages of the logarithmic price ratios $y_n^t - y_{tj} = \ln(p_n^t/p_{tj})$.

[65]This step can be justified as follows. First note that equations (C5) can be replaced by the equivalent equations: (C5)* $\sum_{n=1}^{N} s_n^t\alpha_t + \sum_{n=1}^{N} s_n^t\beta_n = \sum_{n=1}^{N} s_n^t y_n^t$ for $t = 1,\ldots,T$. Now sum equations (C5)* for $t = 1,\ldots,T$ and note that this sum is equal to the sum of equations (C6) for $n = 1,\ldots,N$. Thus the last equation in (C6) is linearly dependent on the remaining equations and can be dropped. This implies that the last equation in (C8) and (C15) can be dropped. Note that if we have found a solution to equations (C5)* and (C6), say $\alpha_1^*,\ldots,\alpha_T^*,\beta_1^*,\ldots,\beta_N^*$, then $\alpha_1^* + \mu,\ldots,\alpha_T^* + \mu, \beta_1^* - \mu,\ldots,\beta_N^* - \mu$ is also a solution to these equations where μ is an arbitrary number. Thus a normalization on these parameters is required in order to obtain a unique solution to the minimization problem (20).

be obtained by solving equations (C15). Denote the solution values to the modified system of equations as $\beta_{1*}, \beta_{2*}, \ldots, \beta_{N1*}$ and let β^* denote the column vector of all of these betas including $\beta_{N*} = 0$. The corresponding parameters b_{n*} which appear in equations (18) are defined as the exponentials of the β_{n*}; i.e., we have:

$$b_{n*} \equiv \exp(\beta_{n*}); \quad n = 1, \ldots, N. \tag{C16}$$

The logarithm of the period t price level, α^{t*}, is then as follows:

$$\alpha^{t*} \equiv \sum_{n=1}^{N} s_n^t (y_n^t - \beta_n^*). \tag{C17}$$

D   The Weighted Time Product Dummy Method: Special Cases

Substituting (23) into (24) and recalling that $y_n^t \equiv ln p_n^t$ and $\beta_n^* = \ln b_n^*$ leads to the following expression for the logarithm of π^t:

$$\ln \pi^t = \sum_{n=1}^{N} s_n^t (y_n^t - \beta_n^*) - \sum_{n=1}^{N} s_n^1 (y_n^1 - \beta_n^*); \quad t = 1, \ldots, T. \tag{D18}$$

It is useful to consider some special cases for the multilateral indexes (D18). Our first special case is the case where we have only two periods, so that $T = 2$. In this case, Diewert (2005; 564) showed that the logarithm of π^2 is defined as follows;

$$\ln \pi^2 \equiv \sum_{n=1}^{N} s_n^* (y_n^2 - y_n^1) \tag{D19}$$

where the share weights s_n^* are defined as the following normalized harmonic averages of the observed expenditure shares s_n^1 and s_n^2 for the two periods:

$$s_n^* \equiv h(s_n^1, s_n^2) / \left[\sum_{k=1}^{N} h(s_k^1, s_k^2) \right]; \quad n = 1, \ldots, N \tag{D20}$$

where $h(s_n^1, s_n^2) \equiv [0.5(s_n^1)^{-1} + 0.5(s_n^2)^{-1}]^{-1}$. Diewert (2005; 564) noted that the bilateral price index defined by π^2 approximated the corresponding Törnqvist index to the second order around a point where $p^1 = p^2$ and $q^1 = q^2$.[66] Thus it may appear that if there are only two periods in the window, the weighted time product dummy (WTPD) multilateral method is a perfectly satisfactory method from the viewpoint of the economic approach to index number theory. Unfortunately, second order approximation results may not be adequate in our present context where a sale for a product. Consider the following example

---

[66]This second order approximation property will hold if $p^2 = \lambda p^1$ and $q^2 = \mu q^1$ where λ and μ are arbitrary positive scalars.

where there are only two commodities and two periods so that $N = 2$ and $T = 2$. Suppose that the period 1 and 2 price vectors are equal to $p^1 \equiv [1, 1]$ and $p^2 \equiv [1, p_{22}]$ and the period 1 and 2 share vectors are equal to $s^1 \equiv [0.5, 0.5]$ and $s^2 \equiv [0, 1]$, where p_{22} is a small positive number that is less than 1. Thus the logarithm of p_{22} is less than zero; i.e., $r \equiv \ln p_{22} < 0$. The price of commodity 2 in period 2 is so low that purchasers stop buying commodity 1 in period 2 even though the price of commodity 1 has not changed. Using formula (5) for the Törnqvist index between periods 2 and 1, the logarithm of this index is defined as follows:

$$
\begin{aligned}
\ln P_T(2/1) & \equiv \sum_{n=1}^{2} (1/2)(s_{1n} + s_{2n}) \ln(p_{2n}/p_{1n}) \\
& = (1/2)[(1/2) + 0] \ln(1/1) + (1/2)[(1/2) + 1] \ln(p_{22}/1) \\
& = (3/4)r
\end{aligned}
\tag{D21}
$$

where $r \equiv \ln p_{22}$ and is negative by assumption. Under our assumptions, $h(s_1^1, s_1^2) = h(0.5, 0) = 0$ and $h(s_2^1, s_2^2) = h(0.5, 1) = 2(0.5)1/[0.5 + 1] = 2/3$. Thus the sum $h(s_1^1, s_1^2) + h(s_2^1, s_2^2) = h(s_2^1, s_2^2) = 2/3$ and substituting these numbers into formula (D19) shows that the logarithm of the WTPD normalized price level for period 2 is equal to the following expression:

$$
\begin{aligned}
\ln \pi^2 & \equiv \sum_{n=1}^{2} s_n^*(y_{2n} - y_{1n}) = 0 \ln(1/1) + 1 \ln(p_{22}/1) \\
& = r < (3/4)r = \ln P_T(2/1)
\end{aligned}
\tag{D22}
$$

where the inequality follows from $r < 0$ by assumption, and the last equality follows using (D21) for our example. This is only a specific numerical example but it shows that price levels based on the weighted time product dummy method may have a substantial bias as compared to the price levels generated by a superlative index.

The second special case that we will consider is the case where the expenditure shares are constant in each period. This case is consistent with purchasers having Cobb-Douglas preferences for the N commodities under consideration. Thus assume that the data satisfy the following conditions:

$$
s_n^t = s_n; \quad t = 1, \ldots, T; n = 1, \ldots, N.
\tag{D23}
$$

Substituting conditions (D23) into (D18) leads to the following formula for the logarithms of the normalized WTPD price levels:

$$
\ln \pi^t = \sum_{n=1}^{N} s_n(y_n^t - y_{1n}); \quad t = 1, \ldots, T.
\tag{D24}
$$

Thus $\pi^t = \prod_{n=1}^{N}(p_n^t/p_{1n})^{s_n}$, a weighted geometric mean of the ratios of the period t prices to the period 1 prices, p_n^t/p_{1n}, using the constant expenditure shares, s_n, as weights.

The third special case we consider is the case where the error terms ε_n^t in equations (19) are all equal to 0. In this case, equations (18) become the following equations:

$$p_n^t = a_t b_n; \quad t = 1, \ldots, T; n = 1, \ldots, N. \tag{D25}$$

Thus in this special case, the period t price vector p^t is equal to $a_t b$, where b is the vector $[b_1, \ldots, b_N]$ and thus the price vectors are proportional to each other. As usual, we can interpret the scalar a_t as the (unnormalized) *period t price level*. The corresponding *period t quantity level*, Q_t, can be defined as period t value, $\sum_{n=1}^{N} p_n^t q_n^t$, divided by the price level, a_t. Thus we have:

$$
\begin{aligned}
Q_t &\equiv p^t \cdot q^t / a_t; \quad t = 1, \ldots, T \\
&= a_t b \cdot q^t / a_t \quad \text{using (D25)} \\
&= b \cdot q^t. \tag{D26}
\end{aligned}
$$

Thus if the error terms in equations (19) are all equal to 0, the period t quantity aggregate, Q_t, is equal to the linear function $b \cdot q^t = \sum_{n=1}^{N} b_n q_n^t$ for all t; i.e., Q_t is equal to a quality adjusted *sum* of the period t quantities purchased. Thus in the case of zero errors in equations (19), the WTPD multilateral method is consistent with purchasers of the N commodities maximizing a linear utility function, which implicitly assumes that the N commodities are perfectly substitutable after quality adjustment. A multilateral method which has this property is called an *additive method*. In the general case where there are nonzero error terms in equations (19), it can be seen that the WTPD multilateral method is an *approximately additive method*.

The problem with an additive multilateral method is that, in general, it is not consistent with the economic approach to index number theory once the number of periods is greater than two; see Marris (1984; 52) and Diewert (1999; 49).

E The Weighted Time Product Dummy Method and the Linking Problem

Suppose that the WTPD multilateral method explained in section 3.3 is used to construct a sequence of price levels, $\pi^1, \pi^2, \ldots, \pi^T$ for T consecutive periods where $T \geq 3$. Recall that the logarithms of these price levels can be defined by equations (D18), $\ln \pi^t = \alpha_t^* - \alpha_{1*} = \sum_{n=1}^{N} s_n^t (y_n^t - \beta_n^*) - \sum_{n=1}^{N} s_{1n}(y_{1n} - \beta_n^*)$ for $t = 1, \ldots, T$, where the $\alpha_{1*}, \alpha_{2*}, \ldots, \alpha_{T*}$ and $\beta_{1*}, \ldots, \beta_N^*$ solved the weighted least squares minimization problem (20) with the normalization $\beta_N^* = 0$.

Assume that data on period $T + 1$ expenditure shares s_n^{T+1} and prices p_n^{T+1} become available at the end of period $T + 1$ for $n = 1, \ldots, N$. A new set of multilateral indexes can now be constructed for the window of observations that include the data for periods $2, 3, \ldots, T, T + 1$. The weighted least squares problem that is solved to generate the price

levels for the new window of observations is the following one:

$$\min_{\alpha^2,\dots,\alpha^{T+1},\beta_1,\dots,\beta_N} \sum_{t=2}^{T+1}\sum_{n=1}^{N} s_n^t(y_n^t - \alpha^t - \beta_n)^2. \tag{E27}$$

Denote the solution to (E27) as $\alpha^{2**},\alpha^{3**},\dots,\alpha^{T+1**},\beta_1^{**},\dots,\beta_N^{**}$ with the normalization $\beta_N^{**} = 0$. The (unnormalized) logarithms of the price levels for the new window are defined as follows:

$$\alpha^{t**} \equiv \sum_{n=1}^{N} s_n^t(y_n^t - \beta_n^{**}); \quad t = 2,3,\dots,T+1. \tag{E28}$$

Recall that $y_n^t = \ln p_n^t$ and $\beta_n^{**} = \ln(b_n^{**})$ for $t = 2,3,\dots,T+1$ and $n = 1,\dots,N$. The period t (unnormalized) price levels a^{t**} are defined as the exponentials of the α^{t**}:

$$a^{t**} \equiv e^{\alpha^{t**}}; \quad t = 2,3,\dots,T+1. \tag{E29}$$

The new WTPD price levels need to be linked to the price levels generated by the original window of observations; i.e., we need to define a new price level for period $T+1$ that extends the initial T normalized price levels π^t defined by equations (D18) for $t = 1,\dots,T$. Suppose we link the new window price levels to the initial price levels at observation t where t could be any observation between periods 2 and T. Let $\pi^{T+1}(t) \equiv \pi^t[a^{T+1**}/a^{t**}]$ denote the resulting (normalized) linked price level for period $T+1$ that depends on choosing period t as link period. Thus the logarithm of this period $T+1$ price level is defined as follows:[67]

$$
\begin{aligned}
\ln\pi^{T+1}(t) &\equiv \ln\pi^t - \ln a^{t**} + \ln a^{T+1**}; \quad t = 2,3,\dots,T \\
&= \sum_{n=1}^{N} s_n^t(y_n^t - \beta_n^*) - \sum_{n=1}^{N} s_n^1(y_n^1 - \beta_n^*) - \alpha^{t**} + \alpha^{T+1**} \quad \text{using (D18) and (E29)} \\
&= \sum_{n=1}^{N} s_n^t(y_n^t - \beta_n^*) - \sum_{n=1}^{N} s_{1n}(y_n^1 - \beta_n^*) - \sum_{n=1}^{N} s_n^t(y_n^t - \beta_n^{**}) \\
&\quad + \sum_{n=1}^{N} s_n^{T+1}(y_n^{T+1} - \beta_n^{**}) \quad \text{using (E28)} \\
&= \sum_{n=1}^{N} s_n^{T+1}(y_n^{T+1} - \beta_n^{**}) - \sum_{n=1}^{N} s_n^1(y_n^1 - \beta_n^*) + \sum_{n=1}^{N} s_n^t(\beta_n^{**} - \beta_n^*). \tag{E30}
\end{aligned}
$$

Thus the choice of the linking observation t between the two windows will only affect the last term in (E30); i.e., the term $\sum_{n=1}^{N} s_n^t(\beta_{n**} - \beta_n^*)$ will in general change as t changes because the period t expenditure shares s_n^t will usually change as t changes. However, as the window length T increases, it is possible that the quality adjustment parameters will stabilize so that $\beta_{n**} \approx \beta_n^*$ for $n = 1,\dots,N$.[68] In this case, the last term on the right

---

[67]de Haan (2015b) essentially obtained the formula (E30).

[68]This observation is likely to be true in a stable market situation where the number of varieties remains

hand side of (E30) will be close to 0 and it will not matter much which period is chosen as the linking period. It can also be seen that in the case where the expenditure shares are constant for all $T + 1$ periods, (E30) simplifies to the following *weighted Jevons formula*:

$$\ln \pi^{T+1}(t) = \sum_{n=1}^{N} s_n(y_n^{T+1} - y_{1n}); \quad t = 2, 3, \ldots, T \quad (E31)$$

where s_n is the constant share for commodity n for $n = 1, \ldots, N$. Thus in the case of constant expenditure shares, the WTPD price level estimate for period $T + 1$ will be independent of the choice of the link period.

In the general case where the shares are not constant, it can be seen that we are faced with the same set of problems that were discussed in section 3.5: we could choose period 2, T or $(T+1)/2$ (if T is odd) as the linking period or we could take the geometric average of the $T - 1$ estimates for π^{T+1} defined by (E30).[69] In the context of very variable price data with little or no trends, choosing the mean estimator for π^{T+1} is probably a satisfactory linking strategy.

Finally, suppose the price and quantity data for period $T + 1$ are identical to the price and quantity data for some period t such that $2 \leq t \leq T$; i.e., we have t such that:

$$s_n^{T+1} = s_n^t; \quad p_n^{T+1} = p_n^t; \quad n = 1, \ldots, N. \quad (E32)$$

Suppose further that we link the two windows at this observation t. The identity test for multilateral indexes requires that $\pi^{T+1}(t) = \pi^t$ where $\pi^{T+1}(t)$ is defined by (E30) and π^t is defined by (D18). Using equation (E30) for period t, we calculate $\ln \pi^{T+1}(t)$ minus $\ln \pi^t$ as follows:

$$
\begin{aligned}
\ln \pi^{T+1}(t) - \ln \pi^t &= \ln a^{T+1^{**}} - \ln a^{t^{**}} \\
&= \alpha^{T+1^{**}} - \alpha^{t^{**}} \\
&= \sum_{n=1}^{N} s_n^{T+1}(y_n^{T+1} - \beta_{n^{**}}) - \sum_{n=1}^{N} s_n^t(y_n^t - \beta_{n^{**}}) \quad \text{using (E28)} \\
&= 0 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (E33)
\end{aligned}
$$

constant. However, this situation is not likely to occur in practice due to the appearance of new items and the disappearance of existing items; i.e., up to now, we have neglected the new goods problem. Thus if the window length T is chosen to be very large, the lack of matching items becomes a problem. Our methodology up to this point has assumed that prices p_n^t and shares s_n^t are all positive. It can be shown that our analysis of the various multilateral methods remains valid when some shares s_n^t are equal to zero provided that the corresponding reservation prices p_n^t remain positive; see Hicks (1940; 114) on the concept of reservation prices. Hausman (1997) estimated these reservation prices econometrically for new breakfast cereals. For a diagram explaining the Hicks methodology, see Diewert (1996; 32). However, the estimation of reservation prices is not a practical alternative for statistical agencies. Feenstra's (1994) CES methodology avoids the estimation of reservation prices since his methodology assumes that they are infinite but in general, an estimate for the elasticity of substitution is required. A more complete treatment of the problems associated with modeling quality change is beyond the scope of this paper.

[69]See definition (33) which also applies in the present context but using the new definitions for the π^t.

where the last equation follows using assumptions (E32). Thus if the price and quantity data for period $T+1$ are exactly equal to the price and quantity data for period t, then linking the windows at observation t will preserve the identity test over the two windows; i.e., under these conditions it makes sense to link the two windows at period t. This logic carries over to situations where the price and quantity data of period t are closest (in some metric) to the price and quantity data of period $T+1$: it makes sense to link the windows at the observation which has the most "similar" price and quantity data.[70] The main advantage of this method is that it will preserve a Walsh type multiperiod identity test across the two windows. The main drawback to this method of linking is that it depends on a *relative price and expenditure share similarity measure* and it may be difficult to achieve consensus on exactly what the appropriate measure is.

F An Efficient Method for Calculating the Geary-Khamis Multilateral Index

Following Diewert (1999; 26), substitute equations (27) into equations (26) and after some simplification, obtain the following system of equations which will determine the components of the b vector:

$$[I_N - C]b = 0_N \tag{F34}$$

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:

$$C \equiv \hat{q}^{-1} \sum_{t=1}^{T} s^t q^{t\mathbf{T}} \tag{F35}$$

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 and $q^{t\mathbf{T}}$ is the transpose of the column vector q^t and is equal to the row vector of quantities purchased during period t. Note that the N by N matrix $s^t q^{t\mathbf{T}}$ has rank 1 for each t. Let $q^{\mathbf{T}}$ denote the transpose of the column vector

---

[70]As in section 3.5, the exact result (E33) can be extended to the case where the price data of period $T+1$ is proportional to the price data of period t.

q. Then:

$$
\begin{aligned}
q^{\mathbf{T}}C &= q^{\mathbf{T}}\hat{q}^{-1}\sum_{t=1}^{T}s^t q^{t\mathbf{T}} \\
&= 1_N{}^{\mathbf{T}}\sum_{t=1}^{T}s^t q^{t\mathbf{T}} \\
&= \sum_{t=1}^{T}1_N{}^{\mathbf{T}}s^t q^{t\mathbf{T}} \\
&= \sum_{t=1}^{T}q^{t\mathbf{T}} \quad \text{since } 1_N{}^{\mathbf{T}}s^t = 1 \text{ for each } t \\
&= q^{\mathbf{T}} \quad \text{using definition (25).} \quad\quad\quad\text{(F36)}
\end{aligned}
$$

Using (F36), we have $q^{\mathbf{T}}[I_N - C] = q^{\mathbf{T}} - q^{\mathbf{T}} = 0_N$. Thus the matrix $I_N - C$ is singular which implies that the N equations in (F34) are not all independent. In particular, if the first N-1 equations in (F34) are satisfied, then the last equation in (F34) will also be satisfied. It can also be seen that the N equations in (F34) are homogeneous of degree one in the components of the vector b. Thus to obtain a unique b solution to (F34), set b_N equal to 1, drop the last equation in (F34) and solve the remaining $N - 1$ equations for $b_1, b_2, \ldots, b_{N-1}$. This is the solution method we used in subsequent sections of this study.[71]

Once the vector of quality adjustment factors b has been determined, the period t (unnormalized) price and quantity levels, P^t and Q_t, are defined using equations (27) and (28); i.e., we have $P^t \equiv p^t \cdot q^t / b \cdot q^t$ and $Q_t \equiv b \cdot q^t$ for each period t in the window. As usual, the sequence of normalized price levels can be defined by $\pi^t \equiv P^t / P^1$ for $t = 1, \ldots, T$.

Equations (28) show that the GK multilateral indexes will be consistent with the economic approach to index number theory provided that purchasers maximize the linear utility function, $f(q) \equiv b \cdot q = \sum_{n=1}^{N} b_n q_n$. However, perhaps somewhat surprisingly, the GK indexes will also be consistent with the economic approach to index number theory if purchasers have Leontief or no substitution preferences.[72] Thus suppose that purchasers of the N products have preferences which can be represented by the utility function f defined as follows:

$$
f(q_1, q_2, \ldots, q_N) \equiv \min\{q_n/\beta_n : n = 1, \ldots, N\} \quad\quad\quad\text{(F37)}
$$

where the β_n are positive constants. Let $\beta \equiv [\beta_1, .., \beta_N]$ the vector of these coefficients. The unit cost function that is dual to these preferences, $c(p)$, turns out to be $c(p) \equiv \beta \cdot p$ where p is the vector of prices that purchasers face. If purchasers in period t face the price vector

---

[71]In order to establish that the solution b_n are all positive, use the argument in Diewert (1999; 27) which proceeds as follows. From (F36), q is a strictly positive solution to $q^{\mathbf{T}}C = q^{\mathbf{T}}I_N = 1q^{\mathbf{T}}$ and hence q is a strictly positive left eigenvector of the strictly positive matrix C which corresponds to the positive eigenvalue 1. Hence by the Perron (1907) Frobenius (1909) Theorem, $\lambda = 1$ is the maximal eigenvalue of C and thus C also has a strictly positive right eigenvector b which corresponds to this positive eigenvalue. Hence b satisfies the equation $Cb = 1b = I_N b$ or $[I_N - C]b = 0_N$ which is (F34).

[72]Diewert (1999; 27) showed that linear and Leontief preferences are the only preferences for which the GK system is exact.

p^t and chose the quantity vector q^t which maximizes the utility function $f(q)$ defined by (F37) so that period t purchaser utility is $u_t \equiv f(q^t)$, then it can be shown that p^t, q^t and u_t satisfy the following equations:

$$q^t = \nabla_p c(p^t) u_t = \beta u_t; \quad t = 1, \ldots, T \tag{F38}$$

where $\nabla_p c(p^t)$ is the vector of first order partial derivatives of the unit cost function $c(p^t) \equiv \sum_{n=1}^{N} \beta_n p_n^t$ and u_t is the period t utility level. Note that equations (F38) imply that the quantity vectors vary in strict proportion over the T time periods. Recall (25) where the vector of total commodity sales over the T periods, $q \equiv \sum_{t=1}^{T} q^t$, was defined. It can be seen that conditions (F38) imply that period t's share of commodity n in total sales of commodity n over the T periods, say $\sigma_n^t \equiv q_n^t / q_n$, does not depend on n. Thus we have the existence of shares $\sigma^t = \sigma_n^t$ for $n = 1, \ldots, N$ and $t = 1, \ldots, T$ such that period t's quantity vector q^t is equal to σ^t times the vector of total sales q; i.e., we have:

$$q^t = \sigma^t q; \quad t = 1, \ldots, T \tag{F39}$$

$$1 = \sum_{t=1}^{T} \sigma^t. \tag{F40}$$

We will use equations (F39) and (F40) to exhibit an explicit solution to equation (F34). Equation (F34) can be rewritten as follows:

$$
\begin{aligned}
b &= Cb \\
&= \hat{q}^{-1} \sum_{t=1}^{T} s^t q^{t\mathbf{T}} b \quad \text{using definition (F35) for C} \\
&= \hat{q}^{-1} \sum_{t=1}^{T} s^t \sigma^t q^{\mathbf{T}} b \quad \text{using (F39)} \\
&= \hat{q}^{-1} \sum_{t=1}^{T} \sigma^t s^t q \cdot b. \tag{F41}
\end{aligned}
$$

We are allowed one normalization on the components of the vector b. We choose the following normalization:

$$b \cdot q = 1. \tag{F42}$$

Equation n in the matrix equation (F41) can be rewritten as follows if we make use of the normalization (F42):

$$b_n \equiv q_n^{-1} \sum_{t=1}^{T} \sigma^t s_n^t; \quad n = 1, \ldots, N. \tag{F43}$$

We need to check that the b_n defined by (F43) satisfy (F42); i.e., we need to show that

$q \cdot b = q^{\mathbf{T}} b = 1$. Using definitions (F43), we have:

$$
\begin{aligned}
q^{\mathbf{T}} b &= 1_N{}^{\mathbf{T}} \sum_{t=1}^{T} \sigma^t s^t \\
&= \sum_{t=1}^{T} \sigma^t 1_N{}^{\mathbf{T}} s^t \\
&= \sum_{t=1}^{T} \sigma^t \quad \text{since } 1_N{}^{\mathbf{T}} s^t = \sum_{n=1}^{N} s_n^t = 1 \text{ for } t = 1, \dots, T \\
&= 1 \quad \text{using (F40).}
\end{aligned}
\tag{F44}
$$

Thus the b_n defined by (F43) satisfy equations (26) subject to the normalization (F42) and these b_n can be used to define the period t (unnormalized) price levels P^t using equations (27); i.e., define $P^t \equiv p^t \cdot q^t / b \cdot q^t$ for $t = 1, \dots, T$. Finally, the sequence of normalized price levels can be defined by $\pi^t \equiv P^t / P^1$ for $t = 1, \dots, T$. Thus the GK price levels are consistent with purchasers having either linear or Leontief preferences over their purchases of the N commodities.

G   Details of Similarity Linking Application

Using the sales adjusted data data, the dissimilarity measures $\triangle_{AL}(p^r, p^t, q^r, q^t)$ defined by (38) using the Fisher bilateral price index as our $P(p^r, p^t, q^r, q^t)$ are calculated for $r = 1, \dots, 12$ and for $t = 1, \dots, 12$. The $\triangle_{AL}(p^1, p^t, q^1, q^t)$ are listed in the column of Table A3 under the heading $\triangle_{AL1t}$, the $\triangle_{AL}(p^2, p^t, q^2, q^t)$ are listed in Table A3 under the heading $\triangle_{AL}^{2t}$, and so on. Note that when $r = t$, $\triangle_{AL}^{tt} = 0$ for each t as expected. Note also that the matrix of measures listed in Table A3 is symmetric so that $\triangle_{AL}^{rt} = \triangle_{AL}^{tr}$ for all r and t.

For each representative elasticity of substitution that we have chosen, we will construct relative price similarity linked normalized price levels, $\pi_{AL}^t(\sigma)$. As each new observation becomes available, we link the current observation to the most similar previous observation. Thus for $\sigma = 0$, we use the measures listed in Table A3. We set $\pi_{AL}^1(0)$ equal to 1. For period 2, we set $\pi_{AL}^2(0) \equiv P_F(p^1, p^2, q^1, q^2)$, the Fisher price index for period 2 relative to period 1. For period 3, we look at the first two entries in the column in Table A3 that has the heading $\triangle_{AL3t}$. The smaller of these two numbers, 0.027, occurs at $t = 1$, so we use the Fisher formula to link the prices of period 3 directly to the prices of period 1; i.e., we set $\pi_{AL}^3(0) \equiv P_F(p^1, p^3, q^1, q^3)$. For period 4, we look at the first three entries in the column in Table A3 that has the heading $\triangle_{AL4t}$. The smallest of these numbers, 0.004, occurs at $t = 3$, so we link the period 4 price level to the price level of period 3 using the Fisher price index for period 4 relative to period 3. Hence the period 4 normalized price level is $\pi_{AL}^4(0) \equiv P_F(p^3, p^4, q^3, q^4) \pi_{AL}^3(0)$. For period 5, we look at the first 4 entries in the column in Table A3 that has the heading $\triangle_{AL5t}$. The smallest of these numbers, 0.074, occurs at $t = 4$, so we link the period 5 price level to the price level of period 4 using the

Table A3: Asymptotic Linear Measures of Price Dissimilarity for $\sigma = 0$

t	$\triangle_{AL}^{1t}$	$\triangle_{AL}^{2t}$	$\triangle_{AL}^{3t}$	$\triangle_{AL}^{4t}$	$\triangle_{AL}^{5t}$	$\triangle_{AL}^{6t}$	$\triangle_{AL}^{7t}$	$\triangle_{AL}^{8t}$	$\triangle_{AL}^{9t}$	$\triangle_{AL}^{10t}$	$\triangle_{AL}^{11t}$	$\triangle_{AL}^{12t}$
1	0.000	0.064	0.027	0.051	0.179	0.102	0.099	0.096	0.130	0.255	0.180	0.221
2	0.064	0.000	0.070	0.119	0.282	0.166	0.184	0.053	0.187	0.384	0.288	0.324
3	0.027	0.070	0.000	0.004	0.105	0.025	0.025	0.042	0.039	0.140	0.068	0.096
4	0.051	0.119	0.004	0.000	0.074	0.008	0.009	0.046	0.018	0.096	0.038	0.058
5	0.179	0.282	0.105	0.074	0.000	0.040	0.048	0.126	0.049	0.016	0.049	0.058
6	0.102	0.166	0.025	0.008	0.040	0.000	0.001	0.049	0.004	0.046	0.015	0.025
7	0.099	0.184	0.025	0.009	0.048	0.001	0.000	0.064	0.002	0.052	0.011	0.022
8	0.096	0.053	0.042	0.046	0.126	0.049	0.064	0.000	0.064	0.170	0.112	0.126
9	0.130	0.187	0.039	0.018	0.049	0.004	0.002	0.064	0.000	0.043	0.004	0.013
10	0.255	0.384	0.140	0.096	0.016	0.046	0.052	0.170	0.043	0.000	0.026	0.027
11	0.180	0.288	0.068	0.038	0.049	0.015	0.011	0.112	0.004	0.026	0.000	0.003
12	0.221	0.324	0.096	0.058	0.058	0.025	0.022	0.126	0.013	0.027	0.003	0.000

60

Table A4: Log Quadratic Measures of Price Dissimilarity for $\sigma = 0$

t	$\triangle_{LQ}^{1t}$	$\triangle_{LQ}^{2t}$	$\triangle_{LQ}^{3t}$	$\triangle_{LQ}^{4t}$	$\triangle_{LQ}^{5t}$	$\triangle_{LQ}^{6t}$	$\triangle_{LQ}^{7t}$	$\triangle_{LQ}^{8t}$	$\triangle_{LQ}^{9t}$	$\triangle_{LQ}^{10t}$	$\triangle_{LQ}^{11t}$	$\triangle_{LQ}^{12t}$
1	0.000	0.063	0.027	0.050	0.174	0.101	0.098	0.095	0.128	0.248	0.177	0.217
2	0.063	0.000	0.068	0.116	0.268	0.161	0.178	0.052	0.180	0.364	0.276	0.311
3	0.027	0.068	0.000	0.004	0.102	0.025	0.025	0.041	0.039	0.137	0.068	0.096
4	0.050	0.116	0.004	0.000	0.071	0.008	0.009	0.045	0.018	0.094	0.038	0.058
5	0.174	0.268	0.102	0.071	0.000	0.039	0.047	0.122	0.048	0.016	0.048	0.057
6	0.101	0.161	0.025	0.008	0.039	0.000	0.001	0.049	0.004	0.045	0.014	0.025
7	0.098	0.178	0.025	0.009	0.047	0.001	0.000	0.063	0.002	0.051	0.011	0.022
8	0.095	0.052	0.041	0.045	0.122	0.049	0.063	0.000	0.063	0.164	0.109	0.123
9	0.128	0.180	0.039	0.018	0.048	0.004	0.002	0.063	0.000	0.042	0.004	0.013
10	0.248	0.364	0.137	0.094	0.016	0.045	0.051	0.164	0.042	0.000	0.026	0.027
11	0.177	0.276	0.068	0.038	0.048	0.014	0.011	0.109	0.004	0.026	0.000	0.003
12	0.217	0.311	0.096	0.058	0.057	0.025	0.022	0.123	0.013	0.027	0.003	0.000

Fisher price index for period 5 relative to period 4. Thus the period 5 normalized price level is $\pi_{AL}^5(0) P_F(p^4, p^5, q^4, q^5) \pi_{AL}^4(0)$. For period 6, we look at the first 5 entries in the column in Table A3 that has the heading $\triangle_{AL6t}$. The smallest of these numbers, 0.008, occurs at $t = 4$, so we link the period 6 price level to the price level of period 4 using the Fisher price index for period 6 relative to period 4. Thus the period 6 normalized price level is $\pi_{AL}^6(0) \equiv P_F(p^4, p^6, q^4, q^6) \pi_{AL}^4(0)$. Continue this sequential process until all 12 of the price similarity linked price levels $\pi_{AL}^t(0)$ have been defined. These price levels are listed in Table A6; see the column with the heading π_{AL}^t.

It can be seen that the price similarity linked price levels π_{AL}^t are quite close to the corresponding target CES price levels, π_{CES}^t but not as close as the CCDI price levels, π_{CCDI}^t, when $\sigma = 0$. Define the mean absolute difference between the π_{CES}^t and the corresponding π_{AL}^t as the average bias, B_{AL} and define the mean absolute differences between the other approximating indexes in a similar manner. Then it turns out that the CCDI and GEKS indexes have the smallest average biases at 0.18 and 0.19 percentage points whereas the price similarity linked indexes π_{AL}^t have an average bias of 0.58 percentage points. The average absolute differences with the π_{CES}^t for the other indexes are listed in Table 4.

We turn now to the Log Quadratic measures of relative price dissimilarity defined by (37). Again, we use the same price and quantity data that were used in the previous two sections. The dissimilarity measures $\triangle_{LQ}(p^r, p^t, q^r, q^t)$ using the Törnqvist bilateral price index as our $P(p^r, p^t, q^r, q^t)$ are calculated for $r = 1, \ldots, 12; t = 1, \ldots, 12$. The resulting dissimilarity measures $\triangle_{LQ}(p^r, p^t, q^r, q^t)$ for $\sigma = 0$ are listed in Table A4. Note that when $r = t$, $\triangle_{LQtt} = 0$ for each t as expected. Note also that the matrix of measures listed in Table A4 is symmetric so that $\triangle_{LQrt} = \triangle_{LQtr}$ for all r and t. Finally, we note that the alternative dissimilarity measures listed in tables A3 and A4 are fairly similar.

For each representative elasticity of substitution σ, we construct relative price similarity linked normalized price levels, $\pi_{LQ}^t(\sigma)$.[73] For $\sigma = 0$, as each new observation becomes available, we link the current observation to the most similar previous observation as was done when we constructed the price levels $\pi_{AL}^t(0)$, except that we now use the dissimilarity measures listed in Table A4 instead of those listed in Table A3. The resulting sequence of price similarity linked price levels $\pi_{LQ}^t(0)$ is listed under the heading π_{LQ}^t in Table A6. It can be seen that the price similarity linked price levels π_{LQ}^t are fairly close to the corresponding target CES price levels, π_{CES}^t, when $\sigma = 0$ with an average absolute difference of 0.62 percentage points. Thus the approximation of the Log Quadratic price similarity linked index $\pi_{LQ}^t(0)$ to the target CES index is similar to the approximation of the Asymptotic Linear price similarity linked index $\pi_{AL}^t(0)$ to π_{CES}^t and both of these approximations are not as good as the approximations provided by the GEKS and CCDI indexes when the elasticity of substitution is 0.

We can construct counterparts to tables A3 and A4 for σ equal to 0.5, 1, 2, 4 and 10

---

[73]As σ changes, the prices do not change but the expenditure shares change. Hence the measures of relative price dissimilarity change as σ changes. We have not listed all of the counterpart tables to tables A3 and A4 but we note that for fixed σ, the AL and LQ dissimilarity measures were quite similar.

and form the price similarity linked indexes $\pi_{AL}^t(\sigma)$ and $\pi_{LQ}^t(\sigma)$ for these values of using this linking methodology. The resulting price levels are listed in Table A6.

H Tables

Table A5: Alternative Price Levels for Different Methods and Elasticities of Substitution

t	π^t_{CES}	π^t_{FCH}	π^t_{TCH}	π^t_{WTPD}	π^t_{GK}	π^t_{GEKS}	π^t_{CCDI}
			Alternative Price Levels when $\sigma = 0$				
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8600	0.8600	0.8585	0.8830	0.8600	0.8600	0.8571
3	0.9500	0.9500	0.9515	0.9227	0.9500	0.9500	0.9507
4	0.9500	0.9500	0.9515	0.9179	0.9500	0.9500	0.9506
5	0.8740	0.8740	0.8726	0.8612	0.8740	0.8740	0.8724
6	0.9700	0.9700	0.9705	0.9367	0.9700	0.9700	0.9705
7	0.9360	0.9360	0.9364	0.9042	0.9360	0.9360	0.9366
8	0.7840	0.7840	0.7829	0.7717	0.7840	0.7840	0.7832
9	0.9100	0.9100	0.9109	0.8816	0.9100	0.9100	0.9106
10	0.8700	0.8700	0.8698	0.8645	0.8700	0.8700	0.8689
11	0.9260	0.9260	0.9268	0.9040	0.9260	0.9260	0.9265
12	0.9300	0.9300	0.9308	0.9144	0.9300	0.9300	0.9303
			Alternative Price Levels when $\sigma = 0.5$				
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8639	0.8626	0.8635	0.8753	0.8523	0.8620	0.8633
3	0.9831	0.9839	0.9834	0.9680	1.0005	0.9831	0.9831
4	0.9915	0.9923	0.9918	0.9735	1.0126	0.9915	0.9915
5	0.8951	0.8935	0.8946	0.8898	0.9046	0.8931	0.8944
6	1.0223	1.0221	1.0223	1.0033	1.0458	1.0221	1.0222
7	0.9831	0.9829	0.9831	0.9650	1.0053	0.9829	0.9831
8	0.8213	0.8200	0.8208	0.8138	0.8314	0.8202	0.8209
9	0.9570	0.9572	0.9571	0.9406	0.9776	0.9568	0.9569
10	0.8990	0.8981	0.8987	0.8974	0.9053	0.8971	0.8983
11	0.9770	0.9770	0.9770	0.9637	0.9948	0.9766	0.9768
12	0.9817	0.9818	0.9817	0.9719	0.9966	0.9811	0.9815
			Alternative Price Levels when $\sigma = 1$				
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8716	0.8675	0.8716	0.8716	0.8529	0.8668	0.8716
3	1.0172	1.0183	1.0172	1.0172	1.0538	1.0160	1.0172
4	1.0346	1.0357	1.0346	1.0347	1.0803	1.0329	1.0346
5	0.9137	0.9078	0.9137	0.9137	0.9329	0.9062	0.9137
6	1.0771	1.0750	1.0771	1.0771	1.1299	1.0744	1.0771
7	1.0321	1.0301	1.0321	1.0322	1.0818	1.0297	1.0321
8	0.8616	0.8567	0.8616	0.8616	0.8857	0.8570	0.8616
9	1.0058	1.0047	1.0058	1.0059	1.0526	1.0033	1.0058
10	0.9244	0.9205	0.9244	0.9244	0.9373	0.9172	0.9244
11	1.0301	1.0285	1.0301	1.0301	1.0719	1.0270	1.0301
12	1.0355	1.0339	1.0355	1.0355	1.0717	1.0316	1.0355
			Alternative Price Levels when $\sigma = 2$				
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8953	0.8801	0.8926	0.8803	0.8776	0.8809	0.8927
3	1.0794	1.0770	1.0790	1.1131	1.1496	1.0696	1.0775
4	1.1151	1.1125	1.1146	1.1588	1.2095	1.1023	1.1126
5	0.9366	0.9156	0.9328	0.9442	0.9774	0.9086	0.9312
6	1.1817	1.1676	1.1791	1.2341	1.3023	1.1643	1.1783
7	1.1248	1.1114	1.1223	1.1738	1.2364	1.1090	1.1217
8	0.9418	0.9207	0.9380	0.9647	1.0013	0.9205	0.9378
9	1.0981	1.0874	1.0961	1.1447	1.2068	1.0818	1.0950
10	0.9559	0.9387	0.9527	0.9558	0.9842	0.9270	0.9502
11	1.1312	1.1186	1.1288	1.1742	1.2348	1.1125	1.1276
12	1.1370	1.1243	1.1346	1.1738	1.2325	1.1153	1.1328

Table A5 (cont.): Alternative Price Levels for Different Methods and Elasticities of Substitution

t	π^t_{CES}	π^t_{FCH}	π^t_{TCH}	π^t_{WTPD}	π^t_{GK}	π^t_{GEKS}	π^t_{CCDI}
Alternative Price Levels when $\sigma = 4$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.9465	0.8924	0.9196	0.9406	0.9670	0.8988	0.9220
3	1.1558	1.1337	1.1447	1.2174	1.2298	1.1177	1.1343
4	1.2209	1.1972	1.2090	1.3145	1.3389	1.1704	1.1925
5	0.9305	0.8881	0.9078	0.9336	0.9771	0.8661	0.8931
6	1.3293	1.2745	1.3000	1.4674	1.5194	1.2610	1.2915
7	1.2529	1.2013	1.2253	1.3772	1.4222	1.1910	1.2185
8	1.0640	0.9890	1.0254	1.1351	1.1815	0.9830	1.0205
9	1.2259	1.1811	1.2017	1.3502	1.4003	1.1627	1.1909
10	0.9453	0.9132	0.9279	0.9297	0.9726	0.8832	0.9088
11	1.2728	1.2244	1.2467	1.3979	1.4545	1.2022	1.2339
12	1.2745	1.2260	1.2484	1.3870	1.4509	1.1954	1.2309
Alternative Price Levels when $\sigma = 10$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0088	0.8565	0.8885	1.0835	1.0912	0.8752	0.9010
3	1.1981	1.1439	1.1568	1.2072	1.2061	1.1266	1.1376
4	1.2933	1.2341	1.2483	1.3174	1.3158	1.1888	1.2046
5	0.8630	0.8710	0.8777	0.8188	0.8293	0.8174	0.8240
6	1.4582	1.3399	1.3687	1.5312	1.5342	1.3042	1.3309
7	1.3631	1.2526	1.2795	1.4270	1.4300	1.2256	1.2490
8	1.1896	0.9935	1.0381	1.3051	1.3141	0.9741	1.0125
9	1.3380	1.2386	1.2629	1.4086	1.4151	1.1962	1.2218
10	0.8636	0.8923	0.8921	0.8165	0.8275	0.8383	0.8429
11	1.3934	1.3216	1.3464	1.4626	1.4726	1.2372	1.2658
12	1.3618	1.2935	1.3172	1.3827	1.3942	1.1997	1.2257
Alternative Price Levels when $\sigma = 20$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0218	0.7971	0.8173	1.0962	1.0962	0.8237	0.8408
3	1.2000	1.1328	1.1411	1.2001	1.2001	1.1249	1.1318
4	1.2998	1.2269	1.2359	1.3011	1.3010	1.1855	1.1949
5	0.8297	0.8727	0.8787	0.8065	0.8107	0.8104	0.8144
6	1.4892	1.3172	1.3387	1.5141	1.5146	1.3048	1.3226
7	1.3907	1.2301	1.2502	1.4129	1.4133	1.2272	1.2427
8	1.2197	0.9157	0.9478	1.3126	1.3125	0.9169	0.9427
9	1.3704	1.2186	1.2376	1.4038	1.4062	1.1960	1.2165
10	0.8297	0.8762	0.8721	0.8065	0.8107	0.8320	0.8349
11	1.4194	1.3645	1.3836	1.4418	1.4461	1.2258	1.2475
12	1.3456	1.3028	1.3198	1.3204	1.3260	1.1755	1.1876

Table A6: Price Levels for the Sales Adjusted Data

t	π^t_{CES}	π^t_{FCH}	π^t_{TCH}	π^t_{FFB}	π^t_{TFB}	π^t_{WTPD}	π^t_{GK}	π^t_{GEKS}	π^t_{CCDI}	π^t_{AL}	π^t_{LQ}
				Alternative Price Levels when $\sigma = 0$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8600	0.8600	0.8585	0.8600	0.8585	0.8828	0.8600	0.8624	0.8612	0.8600	0.8585
3	0.9500	0.9232	0.9169	0.9444	0.9450	0.9296	0.9559	0.9524	0.9531	0.9444	0.9450
4	0.9500	0.9205	0.9141	0.9500	0.9502	0.9179	0.9500	0.9495	0.9505	0.9416	0.9421
5	0.8740	0.8468	0.8384	0.8740	0.8712	0.8615	0.8740	0.8759	0.8751	0.8663	0.8640
6	0.9700	0.9216	0.9091	0.9794	0.9813	0.9357	0.9654	0.9688	0.9691	0.9651	0.9657
7	0.9360	0.8904	0.8784	0.9360	0.9364	0.9042	0.9360	0.9354	0.9364	0.9325	0.9330
8	0.7840	0.7458	0.7343	0.7840	0.7858	0.7715	0.7840	0.7863	0.7863	0.7942	0.7965
9	0.9100	0.8394	0.8224	0.8937	0.8956	0.8659	0.8937	0.9010	0.9008	0.9036	0.9041
10	0.8700	0.7970	0.7789	0.8700	0.8680	0.8648	0.8700	0.8706	0.8703	0.8623	0.8600
11	0.9260	0.8350	0.8140	0.9385	0.9398	0.9103	0.9286	0.9272	0.9279	0.9168	0.9172
12	0.9300	0.8412	0.8201	0.9300	0.9299	0.9144	0.9300	0.9290	0.9297	0.9236	0.9241
				Alternative Price Levels when $\sigma = 0.5$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8639	0.8626	0.8635	0.8626	0.8635	0.8751	0.8527	0.8644	0.8672	0.8626	0.8635
3	0.9831	0.9574	0.9493	0.9790	0.9796	0.9718	1.0020	0.9849	0.9851	0.9790	0.9796
4	0.9915	0.9634	0.9552	0.9915	0.9915	0.9735	1.0126	0.9916	0.9921	0.9852	0.9857
5	0.8951	0.8674	0.8615	0.8919	0.8940	0.8900	0.9045	0.8963	0.8984	0.8870	0.8891
6	1.0223	0.9712	0.9578	1.0354	1.0364	1.0018	1.0392	1.0220	1.0218	1.0206	1.0211
7	0.9831	0.9351	0.9223	0.9830	0.9831	0.9650	1.0053	0.9833	0.9838	0.9827	0.9832
8	0.8213	0.7801	0.7701	0.8224	0.8217	0.8137	0.8317	0.8230	0.8246	0.8334	0.8365
9	0.9570	0.8861	0.8677	0.9427	0.9447	0.9250	0.9606	0.9496	0.9490	0.9539	0.9544
10	0.8990	0.8261	0.8086	0.8960	0.8979	0.8976	0.9052	0.8992	0.9013	0.8907	0.8929
11	0.9770	0.8825	0.8597	0.9923	0.9944	0.9711	0.9964	0.9796	0.9799	0.9719	0.9724
12	0.9817	0.8901	0.8671	0.9806	0.9813	0.9719	0.9965	0.9816	0.9822	0.9802	0.9808
				Alternative Price Levels when $\sigma = 1$							
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8716	0.8675	0.8716	0.8675	0.8716	0.8716	0.8537	0.8691	0.8754	0.8675	0.8716
3	1.0172	0.9929	0.9845	1.0145	1.0154	1.0181	1.0510	1.0170	1.0186	1.0145	1.0154
4	1.0346	1.0083	0.9998	1.0341	1.0346	1.0347	1.0804	1.0338	1.0359	1.0303	1.0311
5	0.9137	0.8838	0.8829	0.9020	0.9137	0.9137	0.9331	0.9106	0.9190	0.9031	0.9106
6	1.0771	1.0225	1.0107	1.0931	1.0950	1.0743	1.1203	1.0756	1.0777	1.0787	1.0795
7	1.0321	0.9810	0.9697	1.0305	1.0321	1.0322	1.0818	1.0312	1.0339	1.0349	1.0357
8	0.8616	0.8159	0.8094	0.8633	0.8616	0.8616	0.8864	0.8605	0.8659	0.8559	0.8578
9	1.0058	0.9345	0.9171	0.9917	0.9960	0.9906	1.0356	0.9982	1.0001	1.0062	1.0069
10	0.9244	0.8512	0.8371	0.9129	0.9244	0.9244	0.9375	0.9208	0.9290	0.9135	0.9213
11	1.0301	0.9318	0.9096	1.0463	1.0524	1.0376	1.0717	1.0321	1.0352	1.0294	1.0302
12	1.0355	0.9407	0.9184	1.0298	1.0355	1.0355	1.0717	1.0336	1.0378	1.0392	1.0402
				Alternative Price Levels when $\sigma = 2$							
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.8953	0.8801	0.8926	0.8801	0.8926	0.8809	0.8802	0.8830	0.8960	0.8801	0.8926
3	1.0794	1.0561	1.0517	1.0782	1.0800	1.0801	1.1395	1.0700	1.0781	1.0782	1.0800
4	1.1151	1.0904	1.0859	1.1111	1.1144	1.1589	1.2102	1.1047	1.1153	1.1133	1.1151
5	0.9366	0.8974	0.9088	0.8920	0.9277	0.9439	0.9793	0.9154	0.9390	0.9163	0.9332
6	1.1817	1.1151	1.1123	1.1983	1.2063	1.2257	1.2850	1.1684	1.1812	1.1885	1.1903
7	1.1248	1.0626	1.0598	1.1142	1.1229	1.1739	1.2373	1.1126	1.1256	1.1325	1.1342
8	0.9418	0.8802	0.8857	0.9382	0.9411	0.9651	1.0039	0.9251	0.9432	0.9356	0.9383
9	1.0981	1.0224	1.0129	1.0763	1.0896	1.1309	1.1913	1.0809	1.0934	1.1039	1.1056
10	0.9559	0.8787	0.8757	0.9094	0.9462	0.9555	0.9861	0.9336	0.9579	0.9347	0.9523
11	1.1312	1.0210	1.0068	1.1404	1.1601	1.1792	1.2290	1.1222	1.1370	1.1378	1.1397
12	1.1370	1.0320	1.0178	1.1085	1.1315	1.1738	1.2336	1.1204	1.1382	1.1500	1.1521

Table A6 (cont.): Price Levels for the Sales Adjusted Data

t	π_{CES}^t	π_{FCH}^t	π_{TCH}^t	π_{FFB}^t	π_{TFB}^t	π_{WTPD}^t	π_{GK}^t	π_{GEKS}^t	π_{CCDI}^t	π_{AL}^t	π_{LQ}^t
					Alternative Price Levels when $\sigma = 4$						
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.9465	0.8924	0.9196	0.8924	0.9196	0.9439	0.9735	0.9006	0.9244	0.8924	0.9196
3	1.1558	1.1236	1.1311	1.1539	1.1560	1.2062	1.2171	1.1193	1.1355	1.1539	1.1560
4	1.2209	1.1868	1.1949	1.2084	1.2152	1.3153	1.3405	1.1750	1.1970	1.2187	1.2212
5	0.9305	0.8804	0.8972	0.8200	0.8608	0.9347	0.9833	0.8758	0.9035	0.9041	0.9170
6	1.3293	1.2284	1.2414	1.3299	1.3486	1.4516	1.4947	1.2698	1.2978	1.3390	1.3419
7	1.2529	1.1584	1.1706	1.2126	1.2337	1.3783	1.4251	1.1979	1.2252	1.2627	1.2653
8	1.0640	0.9537	0.9796	1.0193	1.0411	1.1386	1.1891	0.9891	1.0269	0.9912	1.0281
9	1.2259	1.1309	1.1373	1.1694	1.1979	1.3398	1.3912	1.1681	1.1957	1.2347	1.2373
10	0.9453	0.8726	0.8760	0.8358	0.8737	0.9308	0.9790	0.8932	0.9197	0.9183	0.9315
11	1.2728	1.1328	1.1338	1.2487	1.2877	1.4002	1.4445	1.2198	1.2503	1.2856	1.2885
12	1.2745	1.1433	1.1446	1.1691	1.2180	1.3887	1.4560	1.2050	1.2405	1.2975	1.3007
					Alternative Price Levels when $\sigma = 10$						
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0088	0.8565	0.8885	0.8565	0.8885	1.0846	1.0924	0.8768	0.9025	0.8565	0.8885
3	1.1981	1.1434	1.1562	1.1971	1.1974	1.2061	1.2052	1.1305	1.1410	1.1971	1.1974
4	1.2933	1.2337	1.2477	1.2847	1.2874	1.3176	1.3159	1.1949	1.2100	1.2916	1.2921
5	0.8630	0.8707	0.8773	0.7515	0.7539	0.8210	0.8309	0.8284	0.8352	0.9116	0.9085
6	1.4582	1.3077	1.3280	1.4227	1.4406	1.5206	1.5217	1.3163	1.3394	1.4632	1.4649
7	1.3631	1.2223	1.2412	1.2860	1.3094	1.4278	1.4305	1.2346	1.2574	1.3676	1.3692
8	1.1896	0.9695	1.0070	1.0239	1.0585	1.3065	1.3156	0.9809	1.0190	0.9948	1.0368
9	1.3380	1.2082	1.2246	1.2114	1.2462	1.4088	1.4152	1.2058	1.2308	1.3464	1.3481
10	0.8636	0.8703	0.8650	0.7762	0.7776	0.8187	0.8291	0.8501	0.8549	0.9122	0.9091
11	1.3934	1.2356	1.2468	1.2805	1.3252	1.4768	1.4828	1.2620	1.2892	1.4097	1.4115
12	1.3618	1.2228	1.2336	1.0983	1.1325	1.3857	1.3963	1.2120	1.2376	1.3950	1.3966

Additional References

Diewert, W.E. (1996), "Comments on CPI Biases", *Business Economics* 32:2, 30-35.

Frobenius, G. (1909), "Über Matrizen aus positiven Elementen", *Sitzungberichte der königlich preussischen Akademie der Wissenschaften*, Part 1, 514-518.

Hausman, J. (1997), "Valuation of New Goods Under Perfect and Imperfect Competition", pp. 209-237 in *The Economics of New Goods*, T. Bresnahan and R. Gordon (eds.), NBER Studies in Income and Wealth, Volume 58, Chicago: The University of Chicago Press.

Perron, O. (1907), "Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus", *Mathematische Annalen* 64, 1-76.