

# Decomposing Multilateral Price Indexes into the Contributions of Individual Commodities

*Michael Webster<sup>1</sup> and Rory C. Tarnow-Mordi<sup>2</sup>*

This article describes methods for decomposing price indexes into contributions from individual commodities, to help understand the influence of each commodity on aggregate price index movements.

Previous authors have addressed the decomposition of bilateral price indexes, which aggregate changes in commodity prices from one time period to another. Our focus is the decomposition of multilateral price indexes, which aggregate commodity prices across more than two time periods or countries at once. Multilateral indexes have historically been used for spatial comparisons, and have recently received attention from statistical agencies looking to produce temporal price indexes from large and high frequency price data sets, such as scanner data. Methods for decomposing these indexes are of practical relevance.

We present decompositions of three multilateral price indexes. We also review methods proposed by other researchers for extending multilateral indexes without revising previously published index levels, and show how to decompose the extended indexes they produce. Finally, we use a data set of seasonal prices and quantities to illustrate how these decomposition methods can be used to understand the influence of individual commodities on multilateral price index movements, and to shed light on the relationships between various multilateral and extension methods.

*Key words:* Scanner data; time product dummy; GEKS; Geary-Khamis; linking indexes.

## 1. Decomposition of Bilateral Price Indexes

Price indexes are used to combine the price changes of individual commodities into an aggregate measure of price change. Statistical agencies also find it useful to work in the opposite direction: to decompose a price index into the contributions of individual commodities. This facilitates the identification of the commodities with the greatest contributions to change, which is helpful for validating the inputs and explaining the index (ILO et al. 2004, chap. 9).

It is useful to start with a few straightforward examples. A price index that takes the form of an arithmetic mean of commodity price ratios or relatives has an *additive*

<sup>1</sup> Australian Bureau of Statistics, 45 Benjamin Way, Belconnen, ACT 2617, Australia. Email: michael.webster@abs.gov.au

<sup>2</sup> Exposé: Data Exposed, Margaret Graham Building, Lot 14, Frome Road, Adelaide, SA 5000 Australia. Email: rory.tarnow-mordi@exposedata.com.au

**Acknowledgments:** This research was performed while Rory C. Tarnow-Mordi was employed by the Australian Bureau of Statistics. Views expressed are those of the authors and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted or used, they should be attributed clearly to the authors. The authors would like to thank an Associate Editor, two anonymous referees, Frances Krsinich, Jan de Haan, Lyndon Ang, Marcel van Kints, Justin Farrow, Siu-Ming Tam and Daniel Melser for helpful comments on earlier drafts. We take full responsibility for any errors in the manuscript that remain.

*decomposition*. In other words, it can be decomposed into a sum of contributions, each depending on prices (or price changes) of an individual commodity:

$$P^{0,1} = \sum_i c_i(\mathbf{p}_i, \mathbf{w}_i) \quad (1)$$

where  $\mathbf{p}_i$  is a vector of prices for commodity  $i$ ,  $\mathbf{w}_i$  is a weight (vector) used to aggregate its prices with the prices of other commodities, and  $c_i$  is some unspecified function that depends only on prices and weights of commodity  $i$ . Note that the subscript in  $c_i$  is not strictly necessary but is included to simplify references to summation terms.

For instance, the Laspeyres index between two periods (0 and 1) can be expressed as

$$P_L^{0,1} = \frac{\sum_i p_i^1 q_i^0}{\sum_i p_i^0 q_i^0} = \sum_i s_i^0 \frac{p_i^1}{p_i^0} \quad (2)$$

where  $p_i^0$  and  $p_i^1$  are the prices of commodity  $i$  in periods 0 and 1,  $q_i^0$  is the quantity of commodity  $i$  in period 0, and  $s_i^0 = p_i^0 q_i^0 / \sum_{j=1}^N p_j^0 q_j^0$  is the expenditure share of commodity  $i$  in period 0.

Similarly, an index that can be expressed as a geometric mean of price relatives has a simple *multiplicative decomposition*:

$$P^{0,1} = \prod_i c_i(\mathbf{p}_i, \mathbf{w}_i) \quad (3)$$

For instance, the Törnqvist index between 0 and 1 can be expressed as

$$P_T^{0,1} = \prod_i \left( \frac{p_i^1}{p_i^0} \right)^{\frac{1}{2}(s_i^0 + s_i^1)} \quad (4)$$

where  $s_i^1$  is the expenditure share of commodity  $i$  in period 1.

Several authors have written about the decomposition of common bilateral price indexes. Balk (2008, chap.4) provides a good overview of the topic. As well as presenting decompositions of the straightforward type above — additive decompositions of arithmetic mean indexes, and multiplicative decompositions of geometric mean indexes — Balk also presents additive decompositions of geometric mean indexes, multiplicative decompositions of arithmetic mean indexes, and both arithmetic and multiplicative decompositions of Fisher and Walsh indexes, referencing earlier publications by Van IJzeren (1952, 1983), Vartia (1974, 1976), Diewert (2002) and Reinsdorf et al. (2002).

Many of these decompositions feature a logarithmic mean involving the price index that is being decomposed. For example, Balk shows that a general arithmetic mean index  $P^{0,1} = \sum_i w_i \frac{p_i^1}{p_i^0}$  can also be written as

$$P^{0,1} = \prod_i \left( \frac{p_i^1}{p_i^0} \right)^{\sigma_i} \quad (5)$$

where  $\left(\frac{p_i^1}{p_i^0}\right)^{\sigma_i}$  is the contribution of commodity  $i$  to the arithmetic mean index:

$$\sigma_i = \frac{w_i \times L(P^{0,1}, p_i^1/p_i^0)}{\sum_j w_j \times L(P^{0,1}, p_j^1/p_j^0)}$$

and  $L$  is the logarithmic mean function, defined as  $L(x, y) = \begin{cases} (y - x)/(\ln y - \ln x) & x \neq y \\ x & x = y \end{cases}$  for positive arguments  $x$  and  $y$ .

There are several possible decompositions of a single price index: for example, the Laspeyres index has an additive decomposition given by Equation 2, as well as a multiplicative decomposition given by Equation 5 where  $w_i = s_i^0$ . The commodity contributions from some decompositions, such as Equation 5, depend on the aggregate price changes (or levels): in the remainder of this article, we refer to such decompositions as *reflexive*. We refer to decompositions with commodity contributions that depend only on the prices of the relevant commodity and the expenditures (or quantities) of any or all commodities, such as Equations 2 and 4, as *simple*. This distinction has not previously been named in any source that we are aware.

Different decompositions may be useful in different scenarios. For instance, when we are comparing the properties of two price indexes, it is useful to decompose them in similar ways; when we are combining index movements additively or multiplicatively, a corresponding (additive or multiplicative) decomposition facilitates the calculation of contributions to the combined index.

Fundamentally, however, if we are interested in separating out the contributions of individual commodities to a price index, a simple decomposition seems preferable to a reflexive decomposition. This is because the aggregate price change, which the reflexive decomposition explicitly references, necessarily depends on the prices of all commodities.

It seems unavoidable for the contributions to depend on expenditures (or quantities) as these reflect measures of economic importance that are used to aggregate the price index. For a simple decomposition, what is important is that the expenditures (or quantities) do not depend on the price index.

Note that Equation 4 also yields a simple decomposition into the contributions of individual price observations:

$$P_T^{0,1} = \prod_i \prod_t (p_i^t)^{f_i^t(\mathbf{s})} \quad (6)$$

$$\text{where } f_i^t(\mathbf{s}) = \begin{cases} -1/2(s_i^0 + s_i^1) & t = 0 \\ 1/2(s_i^0 + s_i^1) & t = 1 \end{cases}$$

It can be shown that decompositions of this general form are unique: if, for a given price index formula, there exist functions of expenditure shares  $f_i^t(\mathbf{s})$  and  $g_i^t(\mathbf{s})$  satisfying  $P^{0,1} = \prod_i \prod_t (p_i^t)^{f_i^t(\mathbf{s})} = \prod_i \prod_t (p_i^t)^{g_i^t(\mathbf{s})}$  for any sets of prices  $p_i^t$  and expenditure shares  $s_i^t$ , then  $f_i^t(\mathbf{s}) = g_i^t(\mathbf{s})$  for all  $i$  and  $t$ . We meet other decompositions of this form later in the article.

## 2. Decomposition of Multilateral Price Indexes

The price indexes mentioned in the previous section are *bilateral*, in the sense that they measure price change between two time periods 0 and 1. Suppose we are interested in measuring price change over a *window* of adjacent time periods between 0 and  $T$ , with  $T > 1$ . Traditional practice involves either calculating a sequence of bilateral indexes between 0 and each subsequent period  $\{P^{0,1}, P^{0,2}, \dots, P^{0,T}\}$  or a sequence of bilateral indexes between consecutive periods  $\{P^{0,1}, P^{1,2}, \dots, P^{T-1,T}\}$ . The former sequence yields a *direct* bilateral index and the latter sequence yields a *chained* bilateral index. Alternatively, we can use a *multilateral* index method to simultaneously estimate a system of price comparisons  $\{P^0, \dots, P^T\}$ .

Ivancic et al. (2011) proposed using multilateral methods to produce price indexes from data sets of retail transactions, finding they gave more satisfactory results than either direct or chained bilateral indexes. This has inspired further studies at several statistical agencies with access to scanner data: see, for instance, De Haan and Krsinich (2014), De Haan (2015), Howard et al. (2015), Chessa (2015), Krsinich (2016), Australian Bureau of Statistics (2016, 2017).

A feature of multilateral indexes is that the price comparison between any pair of time periods  $a$  and  $b$  may depend on prices in other periods, and on commodities that are sold in  $a$  and not  $b$  or vice versa. This makes it important to be able to decompose multilateral index movements: without this, it is challenging to interpret which commodities' price changes have the greatest influence on price comparisons.

The decomposition of multilateral price indexes is the focus of the remainder of this paper. We decompose three multilateral methods considered in the studies cited above:

1. The Time Product Dummy (TPD) method advocated by Krsinich (2016), which is a temporal analogue of the Country Product Dummy method introduced by Summers (1973),
2. The GEKS method proposed by Gini (1931), Eltetö and Köves (1964) and Szulc (1964), especially the GEKS-Törnqvist or CCD variant proposed by Caves et al. (1982),
3. The Geary-Khamis (GK) method proposed by Geary (1958) and Khamis (1972).

We focus on these specific multilateral methods because a number of statistical agencies are either researching them or starting to use them for the production of official price indexes.

### 2.1. Decomposition of the TPD Method

Suppose we have a set of price observations  $p_i^t$  pertaining to periods  $t \in \{0, \dots, T\}$  and commodities  $i \in \{1, \dots, N\}$ , possibly with some missingness (combinations of  $i$  and  $t$  for which  $p_i^t$  is not observed or does not exist).

The TPD method involves calculating a system of price comparisons by fitting the model

$$\ln p_i^t = \alpha + \delta^t + \gamma_i + \varepsilon_i^t \quad (7)$$

where  $\alpha$  is the intercept,  $\delta^t$  is the *time effect* parameter for period  $t$ ,  $\gamma_i$  is the *product (commodity) effect* parameter for commodity  $i$  and  $\varepsilon_i^t$  is an error term. In estimating

the model, we choose an arbitrary time period and commodity to treat as reference categories, and set their effects to zero: for notational convenience, we select period 0 and commodity  $N$ .

The remaining parameters in the model are estimated by minimising the sum of squared residuals. Where expenditure information is available, a common approach is to minimise the sum of weighted squared residuals using the expenditure shares  $s_i^t$  as weights (see [Rao 2005](#); [De Haan and Krsinich 2014](#); [Chessa 2015](#); [Krsinich 2016](#); [Australian Bureau of Statistics 2016](#)).

The time effect parameter estimates reflect the natural logarithm of the price level in each period, relative to period 0, so it is natural to estimate the price level in each period by taking the exponential of the time effect estimates. The TPD price comparison between periods  $a$  and  $b$  is thus the ratio of price levels

$$P_{\text{TPD}}^{a,b} = \frac{\exp(\hat{\delta}^b)}{\exp(\hat{\delta}^a)} = \exp(\hat{\delta}^b - \hat{\delta}^a) \quad (8)$$

Strictly the exponential transformation introduces a model bias, which in this context is usually implicitly or explicitly treated as small enough to ignore (see, for instance, [De Haan et al. 2016](#)).

### 2.1.1. Simple TPD Decompositions

We can decompose TPD price comparisons by following the weighted least squares process used to derive the parameters. In general, regression model parameter estimates under the weighted least squares process are given by the product of matrices

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{p} = \mathbf{A} \mathbf{p} \quad (9)$$

In cases where the design matrix  $\mathbf{X}$  and the weight matrix  $\mathbf{W}$  are considered fixed and known, this equation demonstrates that each parameter is a linear combination (represented by matrix  $\mathbf{A}$ ) of the observed variables  $\mathbf{p}$ ; that is  $\beta_i = \sum_j A_{i,j} p_j$ . This fact, combined with the exponential transformation in Equation 8, gives a natural multiplicative decomposition of the price change between two periods.

Specifically, the weighted least squares equation in our case is composed of

- The parameter estimate vector  $\hat{\beta}$  which is  $\begin{bmatrix} \hat{\alpha} & \hat{\delta}^1 & \cdots & \hat{\delta}^T & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{N-1} \end{bmatrix}^T$
- The design matrix  $\mathbf{X}$  corresponding to the parameter vector and price vector, with a simple structure:

$$\left[ \begin{array}{cccc|ccc} 1 & D^1(1) & \cdots & D^T(1) & D_1(1) & \cdots & D_{N-1}(1) \\ 1 & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & D^1(K) & \cdots & D^T(K) & D_1(K) & \cdots & D_{N-1}(K) \end{array} \right] = [\mathbf{X}_\tau | \mathbf{X}_\pi]$$

where  $D^t(k)$  and  $D_i(k)$  are dummy variables with values of 1 if the  $k$ -th price observation pertains to period  $t$  and commodity  $i$  respectively and zero otherwise.  $K$  is the total number of price observations.

- The weight matrix  $\mathbf{W}$ , a diagonal matrix of expenditure shares:  $\text{diag}(s_i^t)$
- The price vector  $\mathbf{p}$ , which contains the log price observations for each commodity-time,  $[\ln p_1 \ \ln p_2 \ \cdots \ \ln p_k \ \cdots \ \ln p_K]^T$

Note that for simplicity, the price vector is indexed with a single variable  $k$  instead of the separate time variable  $t$  and commodity variable  $i$  shown in Equation 7. This difference is superficial: both representations are equivalent, but a single index variable makes the linear algebra simpler.

As only the time effect parameters are needed to estimate TPD comparisons, the weighted least squares solution of Equation 9 can be simplified using the Banachiewicz formula for block matrix inversion (see [Puntanen and Styan 2006](#)):

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

where the block matrices are of appropriate dimensions and  $\mathbf{A}$ ,  $\mathbf{D}$ , and  $(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})$  are invertible. Applying this result, we obtain

$$\hat{\delta}^a = \left[ (\mathbf{W}_\tau - \mathbf{W}_{\tau\pi}\mathbf{W}_\pi^{-1}\mathbf{W}_{\tau\pi}^T)^{-1} (\mathbf{X}_\tau^T - \mathbf{W}_{\tau\pi}\mathbf{W}_\pi^{-1}\mathbf{X}_\pi^T) \mathbf{W} \right]_{a+1} \mathbf{p} = \sum_{k=1}^K w_{a,k} \ln(p_k) \quad (10)$$

where  $\mathbf{W}_\tau = \mathbf{X}_\tau^T \mathbf{W} \mathbf{X}_\tau$ ,  $\mathbf{W}_\pi = \mathbf{X}_\pi^T \mathbf{W} \mathbf{X}_\pi$ ,  $\mathbf{W}_{\tau\pi} = \mathbf{X}_\tau^T \mathbf{W} \mathbf{X}_\pi$ , and the  $a+1$  subscript indicates we take row  $a+1$  of the matrix in the square brackets. Note that when in the proceeding paragraphs, variable  $a$  may be replaced with variable  $b$ , but the analogical formulas apply.

Equation (10) defines the weights  $w_{a,k}$  for  $a > 0$ ; for  $a = 0$  (the reference period) we set  $w_{0,k} = 0$  for every  $k$ , as the corresponding weights from (10) would yield the parameter estimate  $\hat{\alpha}$ .

This simplification is useful for computation, as it limits the size of the matrix required to be inverted to the number of time periods included in the model. This is a particular advantage for TPD methods that aggregate the prices of an arbitrary number of commodities over a window of a fixed size, as it protects the performance of any implementation.

Combining Equations 8 and 10, it follows that a decomposition of the TPD price index in terms of commodity price observations is

$$\begin{aligned} P_{\text{TPD}}^{a,b} &= \prod_{k=1}^K p_k^{w_{b,k} - w_{a,k}} \\ &= \prod_i \prod_t (p_i^t)^{w_{b,k(i,t)} - w_{a,k(i,t)}} \\ &= \prod_i \prod_t c_{\text{TPD},i}^t(a,b) \end{aligned} \quad (11)$$

where  $k(i, t)$  is the observation corresponding to commodity  $i$  and period  $t$ , and  $c_{\text{TPD},i}^t(a,b) = (p_i^t)^{w_{b,k(i,t)} - w_{a,k(i,t)}}$  is the contribution of price observation  $p_i^t$  to the TPD price comparison between periods  $a$  and  $b$ .

We can alternatively decompose TPD price comparisons by deriving the parameters in a manner similar to [Diewert and Fox \(2017\)](#). The weighted sum of squared errors can be written as

$$E = \sum_i \sum_t s_i^t (\log p_i^t - \alpha - \delta^t - \gamma_i)^2 \quad (12)$$

The time and commodity effect parameters that minimize  $E$  satisfy  $\frac{\partial E}{\partial \delta^t} = 0$  and  $\frac{\partial E}{\partial \gamma_i} = 0$ . This yields a pair of equations

$$\hat{\delta}^t = \sum_i s_i^t (\log p_i^t - (\hat{\alpha} + \hat{\gamma}_i)) \quad (13)$$

$$\hat{\alpha} + \hat{\gamma}_i = \frac{\sum_i s_i^t (\log p_i^t - \hat{\delta}^t)}{\sum_i s_i^t} \quad (14)$$

Substituting Equation 14 into Equation 13 to eliminate  $\hat{\alpha}$  and  $\hat{\gamma}_i$  yields

$$\hat{\delta}^t = \sum_u \left( \sum_i \frac{s_i^t s_i^u}{s_i^\Sigma} \right) \hat{\delta}^u + \sum_i s_i^t \left( \log p_i^t - \frac{\sum_u s_i^u \log p_i^u}{s_i^\Sigma} \right) \quad (15)$$

where  $s_i^\Sigma = \sum_t s_i^t$

Equation 15 can be written in vector-matrix form as  $\mathbf{Id} = \mathbf{Md} + \mathbf{b}$ , where  $\mathbf{I}$  is an identity matrix of size  $T + 1$ ,  $\mathbf{d}$  is a vector of the time effect estimates,  $\mathbf{M}$  is a matrix with the element in the  $t$ -th row and  $u$ -th column equal to

$$\sum_i \frac{s_i^t s_i^u}{s_i^\Sigma},$$

and  $\mathbf{b}$  is a vector with the  $t$ -th element equal to

$$\sum_i s_i^t \left( \log p_i^t - \frac{\sum_u s_i^u \log p_i^u}{s_i^\Sigma} \right)$$

The solution to this equation satisfies

$$(\mathbf{M} - \mathbf{I})\mathbf{d} = -\mathbf{b} \quad (16)$$

The matrix  $\mathbf{M} - \mathbf{I}$  is singular so we cannot invert it to solve Equation (16). However, we usually constrain the time effects by setting  $\delta^0 = 0$ . This constraint can be expressed in matrix form as

$$\mathbf{Cd} = \mathbf{0} \quad (17)$$

where  $\mathbf{C}$  is a matrix with all entries in the first column equal to 1 and 0 elsewhere and  $\mathbf{0}$  is a vector of zeroes. [Collier \(1999\)](#) uses a similar technique in a different context (deriving Geary-Khamis indexes). Adding Equations 16 and 17 yields

$$(\mathbf{M} - \mathbf{I} + \mathbf{C})\mathbf{d} = -\mathbf{b} \quad (18)$$

$$\mathbf{d} = -(\mathbf{M} - \mathbf{I} + \mathbf{C})^{-1}\mathbf{b} \quad (19)$$

Let us denote the element in the  $r$ -th row and  $c$ -th column of  $(\mathbf{M} - \mathbf{I} + \mathbf{C})^{-1}$  as  $m_{rc}$ . Then from Equation 19 we can write an arbitrary time effect estimate  $\hat{\delta}^a$  as

$$\hat{\delta}^a = -\sum_t m_{at} \sum_i s_i^t \left( \log p_i^t - \frac{\sum_u s_i^u \log p_i^u}{s_i^\Sigma} \right) = -\sum_i \sum_t s_i^t \log p_i^t \left( m_{at} - \frac{\sum_u s_i^u m_{au}}{s_i^\Sigma} \right) \quad (20)$$

It follows that

$$\begin{aligned} P_{\text{TPD}}^{a,b} &= \exp(\hat{\delta}^b - \hat{\delta}^a) \\ &= \prod_i \prod_t (p_i^t)^{-s_i^t \left( [m_{bt} - m_{at}] - \sum_u s_i^u [m_{bu} - m_{au}] / s_i^\Sigma \right)} \end{aligned} \quad (21)$$

Equations 11 and 21 give apparently distinct, but actually equivalent, formulations of a simple TPD decomposition. This fact is a consequence of the uniqueness of the solution to a full rank weighted least squares problem. It also follows from our earlier observation (from Section 1) that decompositions of this form are unique. Both formulations can be used to explain the impact of individual commodities by combining the relevant terms, that is,  $\prod_i c_{\text{TPD},i}^t(a, b)$  gives the contribution of commodity  $i$  to the price comparison between periods  $a$  and  $b$ .

Despite the equivalence of the two formulations of this decomposition, the former formulation focuses on simplicity in linear algebra, but loses the explicit separation of time and commodity terms in the decomposition, requiring these to be recovered after the decomposition is derived. The latter formulation carefully maintains the separate time and commodity terms, but is more difficult to express in terms of matrix operations. Both of these decompositions will be referred to henceforth as the Simple TPD Decomposition.

### 2.1.2. Reflexive TPD Decomposition

A third decomposition of the TPD can be derived from a multilateral method proposed by Rao (1990), which involves solving a set of equations

$$P_{\text{Rao}}^t = \prod_i \left( \frac{p_i^t}{\pi_i} \right)^{s_i^t} \quad (22)$$

$$\pi_i = \prod_t \left( \frac{p_i^t}{P_{\text{Rao}}^t} \right)^{\sum_u s_i^u} \quad (23)$$

simultaneously for the unknown parameters  $\pi_i$  and  $P_{\text{Rao}}^t$ .  $\pi_i$  can be interpreted as a *reference price* for commodity  $i$ , and we typically impose the condition  $P_{\text{Rao}}^0 = 1$  to obtain a unique solution.

Rao (2005) demonstrates that the system of price comparisons obtained by solving Equations 22 and 23 simultaneously is equivalent to the (weighted) TPD system.



From Equation 22, the TPD price change between two periods  $a$  and  $b$  can be expressed as

$$P_{\text{TPD}}^{a,b} = \prod_i \frac{(p_i^b)^{s_i^b}}{(p_i^a)^{s_i^a}} (\pi_i)^{s_i^a - s_i^b} \quad (24)$$

We need to be careful expressing the price change in this way, because some commodities may only have a price in one of the two periods  $a$  and  $b$ . Where, for instance, a commodity is not sold in period  $a$ , we simply replace the expenditure share  $s_i^a$  with a 0, and consequently replace the exponentiated missing price  $(p_i^a)^{s_i^a}$  with a 1.

Equation 24 has each commodity's contribution to the price change expressed in terms of its weighted prices, as well as the reference price  $\pi_i$ . From Equation 23, we can see that the reference prices depend on the aggregate price levels, which makes the decomposition reflexive. It also has the interesting property that choosing a period other than 0 as the reference would not alter the price comparisons, but could alter the reference prices  $\pi_i$ , and consequently the commodity contributions to those price comparisons. We will continue to refer to this decomposition as the Reflexive TPD Decomposition.

## 2.2. Decomposition of the CCD and GEKS Methods

The GEKS method involves calculating multilateral price comparisons by combining bilateral Fisher indexes:

$$P_{\text{GEKS}}^{a,b} = \prod_i \left( \frac{P_F^{t,b}}{P_F^{t,a}} \right)^{\frac{1}{T+1}} \quad (25)$$

where  $P_F$  is a Fisher price index:

$$P_F^{0,1} = \left[ \left( \frac{\sum_i p_i^1 q_i^0}{\sum_i p_i^0 q_i^0} \right) \left( \frac{\sum_i p_i^1 q_i^1}{\sum_i p_i^0 q_i^1} \right) \right]^{1/2} \quad (26)$$

GEKS-Törnqvist or CCD price comparisons are obtained by replacing the Fisher indexes in Equation 25 with Törnqvist indexes:

$$P_{\text{CCD}}^{a,b} = \prod_i \left( \frac{P_T^{t,b}}{P_T^{t,a}} \right)^{\frac{1}{T+1}} \quad (27)$$

We can easily derive multiplicative decompositions of the CCD index using a multiplicative decomposition of the Törnqvist index. From Equation 4, we know that the Törnqvist index can be written as a product of commodity contributions

$c_{T,i}(0,1) = \left( \frac{p_i^1}{p_i^0} \right)^{\frac{1}{2(s_i^0 + s_i^1)}}$ . Substituting this into Equation 27, we obtain

$$\begin{aligned}
 P_{\text{CCD}}^{a,b} &= \prod_i \prod_t \left( \frac{c_{T,i}(t,b)}{c_{T,i}(t,a)} \right)^{\frac{1}{T+1}} \\
 &= \prod_i \frac{(p_i^b)^{w_i(\bullet,b)}}{(p_i^a)^{w_i(\bullet,a)}} \left[ \prod_t (p_i^t)^{\frac{w_i(t,a)-w_i(t,b)}{T+1}} \right] \\
 &= \prod_i c_{\text{CCD},i}(a,b)
 \end{aligned} \tag{28}$$

where

$w_i(t,a) = \frac{1}{2} \left( \frac{s_i^t}{\sum_{i \in (t \cap a)} s_i^t} + \frac{s_i^a}{\sum_{i \in (t \cap a)} s_i^a} \right)$  is the weight of commodity  $i$  in the Törnqvist price comparison between periods  $t$  and  $a$  (represented by the notation  $i \in (t \cap a)$ )  
 $w_i(\bullet,a) = \frac{1}{T+1} \sum_t w_i(t,a)$  is the average weight across comparisons involving period  $a$   
and  $c_{\text{CCD},i}(a,b) = \frac{(p_i^b)^{w_i(\bullet,b)}}{(p_i^a)^{w_i(\bullet,a)}} \left[ \prod_t (p_i^t)^{\frac{w_i(t,a)-w_i(t,b)}{T+1}} \right]$  is the contribution of commodity  $i$  to the CCD price comparison between periods  $a$  and  $b$ .

Note that this CCD decomposition is simple: it inherits this property from the Törnqvist decomposition.

If there are any missing prices  $p_i^t$ , we replace the corresponding term(s) with a 1. If there are no missing prices (i.e., the same set of commodities is sold every period), Equation 28 can be simplified to an expression based on the CCDI index presented by [Diewert and Fox \(2017\)](#):

$$P_{\text{CCD}}^{a,b} = \prod_i \frac{(p_i^b)^{\frac{1}{2}(s_i^\bullet + s_i^b)}}{(p_i^a)^{\frac{1}{2}(s_i^\bullet + s_i^a)}} (p_i^\bullet)^{\frac{1}{2}(s_i^a - s_i^b)} \tag{29}$$

where  $s_i^\bullet = \frac{1}{T+1} \sum_t s_i^t$  and  $p_i^\bullet = \prod_t (p_i^t)^{\frac{1}{T+1}}$

As observed by [Chessa et al. \(2017\)](#), Equation 29 can be expressed as a geometric average of two factors. The second factor is very similar to Equation 24, revealing that the TPD and CCD indexes are closely related. However, the first factor reveals that the CCD gives more influence to local price changes between periods  $a$  and  $b$ .

$$P_{\text{CCD}}^{a,b} = \prod_i \left[ \left( \frac{p_i^b}{p_i^a} \right)^{s_i^\bullet} \right]^{\frac{1}{2}} \left[ \frac{(p_i^b)^{s_i^b}}{(p_i^a)^{s_i^a}} (p_i^\bullet)^{(s_i^a - s_i^b)} \right]^{\frac{1}{2}} \tag{30}$$

We could obtain a multiplicative decomposition of a GEKS price comparison in a similar way, by substituting a multiplicative Fisher decomposition into Equation 25. The results are not presented here. We note, however, that a GEKS decomposition will inherit the simple/reflexive property of the corresponding Fisher decomposition. The multiplicative Fisher decompositions presented by [Balk \(2008, chap. 4\)](#) are both reflexive.

### 2.3. Decomposition of the GK Method

The GK method involves solving a set of simultaneous equations, similar to those from the Rao method:

$$P_{\text{GK}}^t = \frac{\sum_i p_i^t q_i^t}{\sum_i \pi_i q_i^t} \quad (31)$$

$$\pi_i = \frac{\sum_t p_i^t q_i^t / P_{\text{GK}}^t}{\sum_t q_i^t} \quad (32)$$

where again,  $\pi_i$  can be interpreted as a reference price for commodity  $i$ , and we typically impose the condition  $P_{\text{GK}}^0 = 1$  to obtain a unique solution.

To decompose GK index movements, it is helpful to first rewrite Equation 31 as

$$P_{\text{GK}}^t = \sum_i \sigma_i^t \frac{p_i^t}{\pi_i} \quad (33)$$

where  $\sigma_i^t = \frac{\pi_i q_i^t}{\sum_j \pi_j q_j^t}$  can be interpreted as an expenditure share of commodity  $i$  in period  $t$ , if all commodities were sold at reference prices.

Instinctively, one might seek an additive decomposition of the GK price change between two periods using Equation 33: some algebraic manipulation yields

$$\begin{aligned} P_{\text{GK}}^{a,b} &= \frac{\sum_i \sigma_i^b \frac{p_i^b}{\pi_i}}{\sum_i \sigma_i^a \frac{p_i^a}{\pi_i}} \\ &= \left[ \sum_{q_i^b > 0} s_i^b \frac{\sigma_i^b p_i^b}{\sigma_i^a p_i^a} + (P_{\text{GK}}^b)^{-1} \sum_{q_i^b = 0} \sigma_i^a \frac{p_i^a}{\pi_i} \right]^{-1} \\ &= \sum_{q_i^b > 0} s_i^a \frac{\sigma_i^b p_i^b}{\sigma_i^a p_i^a} + (P_{\text{GK}}^a)^{-1} \sum_{q_i^b = 0} \sigma_i^b \frac{p_i^b}{\pi_i} \end{aligned} \quad (34)$$

where the first sum includes commodities sold in both  $a$  and  $b$ , and the second sum includes commodities sold in  $b$  and not  $a$  (last line of Equation 34) or vice versa (second last line). The last line of Equation 34 is an additive decomposition that is reflexive through its inclusion of the aggregate price level  $P_{\text{GK}}^a$ , and also indirectly through the shares  $\sigma_i^t$  and the reference prices  $\pi_i$ .

When an identical set of commodities is sold in  $a$  and  $b$ , the second term of this decomposition disappears and the first term seems quite appealing as an additive decomposition. However, in general, the asymmetric manner in which it handles prices

missing from only one of  $a$  or  $b$  seems unsatisfactory. Taking the mean of the second last and last lines of Equation 34 would address the asymmetry but the result is no longer an additive decomposition.

We can obtain a more symmetric GK decomposition by first using Equation 5 to convert Equation 33 to a multiplicative form:

$$P_{\text{GK}}^t = \prod_i \left( \frac{p_i^t}{\pi_i} \right)^{\theta_i^t} \quad (35)$$

where

$$\begin{aligned} \theta_i^t &= \frac{\sigma_i^t \times L(P_{\text{GK}}^t, p_i^t / \pi_i)}{\sum_j \sigma_j^t \times L(P_{\text{GK}}^t, p_j^t / \pi_j)} \\ &= \frac{q_i^t \times L(\pi_i P_{\text{GK}}^t, p_i^t)}{\sum_j q_j^t \times L(\pi_j P_{\text{GK}}^t, p_j^t)} \\ &\approx \frac{q_i^t \times p_i^t}{\sum_j q_j^t \times p_j^t} \\ &= s_i^t \end{aligned}$$

where the second equality follows from the definition of  $\sigma_i^t$  and the homogeneity of the logarithmic mean, and the approximation follows from the  $p_i^t \approx \pi_i P_{\text{GK}}^t$  relationship implicit in the GK method.

It follows that

$$\begin{aligned} P_{\text{GK}}^{a,b} &= \frac{\prod_i \left( \frac{p_i^b}{\pi_i} \right)^{\theta_i^b}}{\prod_i \left( \frac{p_i^a}{\pi_i} \right)^{\theta_i^a}} \\ &= \prod_i \frac{(p_i^b)^{\theta_i^b}}{(p_i^a)^{\theta_i^a}} (\pi_i)^{\theta_i^a - \theta_i^b} \end{aligned} \quad (36)$$

Equation 36 is a multiplicative GK decomposition that is reflexive through both the reference prices  $\pi_i$  and the exponents  $\theta_i^t$ . Note the similarity to Equation 24.

### 3. Decomposition of Extended Multilateral Indexes

Statistical agencies compute and publish price indexes as new periods of price data become available. The published index series is extended by linking or “splicing” price comparisons involving the latest period onto published index levels for previous periods. This section focusses on how to extend the index series when multilateral methods are

used to generate price comparisons, and how we can use the results from the previous section to decompose published index movements.

It is relatively straightforward to extend a bilateral index. For instance, when data from period  $t$  becomes available, we would extend a direct bilateral index to period  $t$  by first calculating the price comparison  $P^{0,t}$  between the reference period (0) and  $t$ , and then multiplying this by the index level in the reference period:  $P^t = P^0 \times P^{0,t}$ . Similarly, we would extend a chained bilateral index by multiplying the previous index level  $P^{t-1}$  by the price comparison between the previous and current periods  $P^{t-1,t}$ .

How best to extend a multilateral price index is more ambiguous. In period  $t$ , we simultaneously estimate price comparisons between  $t$  and several historical periods. Using the price comparison from one historical period to extend the index may yield a different result to using the price comparison from another.

Several authors have proposed splicing methods for extending multilateral indexes. In this article, we focus on decomposing the methods considered in [Australian Bureau of Statistics \(2017\)](#):

- *Rolling window* methods, including the *movement splice* proposed by [Ivancic et al. \(2011\)](#), the *window splice* proposed by [Krsinich \(2016\)](#), the *half (window) splice* proposed by [De Haan \(2015\)](#), and the *mean splice* proposed by [Diewert and Fox \(2017\)](#). These methods involve selecting a fixed window length ( $T + 1$  periods) for multilateral comparisons. As each new period of data becomes available, we calculate a new system of comparisons over the window spanning from  $t - T$  to  $t$  and splice it together with the previous system of comparisons (using a window spanning from  $t - T - 1$  to  $t - 1$ ) to estimate the index movement from  $t - 1$  to  $t$ ,
- The *direct* method proposed by [Chessa \(2015\)](#). This method involves selecting a fixed base period  $b$  (say, December) as the start of the multilateral comparison window. As each new period of data becomes available, we calculate a system of comparisons spanning from  $b$  to  $t$  and use the direct price comparison between  $b$  and  $t$  to estimate the price change between these periods. The base period can be updated regularly (e.g., annually).

[Table 1](#) expresses the extended price movements between consecutive periods ( $t - 1$  and  $t$ ) in terms of multilateral price movements from the current window (ending in  $t$ ) and the previous window (ending in  $t - 1$ ). The methods are algebraically similar, though in practice the indexes may yield different trends. The next section presents empirical results.

Of most relevance here is that they all combine multilateral price movements in a multiplicative manner (through division or geometric averaging). This means that we can substitute a multiplicative decomposition for each of the multilateral price movements that feature in the extended price movement, and collect like terms to obtain a multiplicative decomposition of the extended price movement. Importantly, if the multilateral decomposition is simple, then the decomposition of the extended price movement is also simple.

In [Table 1](#),  $P^{x,y}(z)$  denotes the aggregate price comparison between periods  $x$  and  $y$  derived from a multilateral window ending in period  $z$ , and  $c_i(x,y;z)$  denotes the contribution of commodity  $i$  to that aggregate comparison.

Table 1. Comparison of extension methods.

Extension method	Price movement between consecutive periods	Decomposition of consecutive movement
Movement splice	$P^{t-1,t} = P^{t-1,t}(t)$	$P^{t-1,t} = \prod_i c_i(t-1,t;t)$
Window splice	$P^{t-1,t} = \frac{P^{t-T,t}(t)}{P^{t-T,t-1}(t-1)}$	$P^{t-1,t} = \prod_i \frac{c_i(t-T,t;t)}{c_i(t-T,t-1;t-1)}$
Half splice (assuming $T$ is even)	$P^{t-1,t} = \frac{P^{t-T/2,t}(t)}{P^{t-T/2,t-1}(t-1)}$	$P^{t-1,t} = \prod_i \frac{c_i(t-T/2,t;t)}{c_i(t-T/2,t-1;t-1)}$
Mean splice	$P^{t-1,t} = \prod_{s=1}^T \left[ \frac{P^{t-s,t}(t)}{P^{t-s,t-1}(t-1)} \right]^{\frac{1}{T}}$	$P^{t-1,t} = \prod_i \left[ \prod_{s=1}^T \frac{c_i(t-s,t;t)}{c_i(t-s,t-1;t-1)} \right]^{\frac{1}{T}}$
Direct	$P^{t-1,t} = \frac{P^{b,t}(t)}{P^{b,t-1}(t-1)}$	$P^{t-1,t} = \prod_i \frac{c_i(b,t;t)}{c_i(b,t-1;t-1)}$

It may be of interest to decompose longer term (e.g., annual) price comparisons of an extended price index. These longer term movements can be expressed as a product of consecutive price movements:

$$P^{a,b} = \prod_{t=a+1}^b P^{t-1,t}$$

(37)

As above, we can substitute a multiplicative decomposition for each element of Equation 37 and collect like terms to obtain a multiplicative decomposition into commodity contributions

$$P^{a,b} = \prod_i \prod_{t=a+1}^b c_i(t-1,t)$$

where  $c_i(t-1,t)$  is the contribution of commodity  $i$  to the extended movement between  $t-1$  and  $t$  (as given in the third column of Table 1). Once again, if the underlying multilateral decomposition is simple, this will be preserved.

4. Empirical Results

In this section, we illustrate how the decomposition methods described in the previous sections can be used to quantify the contributions of individual commodities to multilateral price comparisons. In Subsection 4.1 we introduce the data used for this analysis. In Subsection 4.2, we decompose indexes calculated using a range of multilateral methods, and in Subsection 4.3 we decompose indexes calculated using a range of extension methods. This allows us to compare and contrast the methods considered. However, in practice, a statistical agency may prefer a single combination of multilateral and extension methods for various reasons. In this context, the comparison between methods is less important than the illustration that we can decompose an index calculated using any combination of the multilateral and extension methods described above.

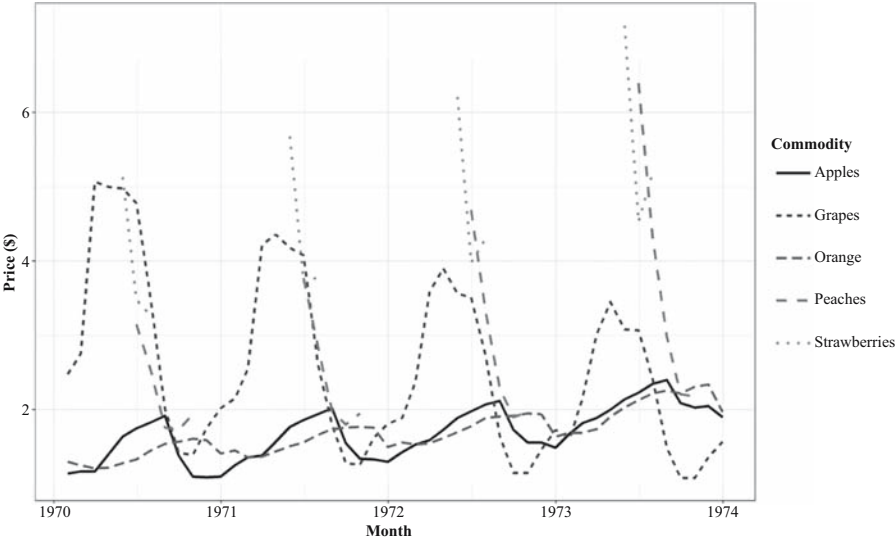


Fig. 1. Price of fruit commodities.

4.1. Data

The main data set we use for this illustration contains monthly price and quantity information relating to five fruit commodities over a period of four years. It is taken from the IWGPS Consumer Price Index Manual (ILO et al. 2004, chap. 22) and is a modified version of a data set from Turvey (1979). Three of the commodities (Apples, Grapes and Oranges) are sold every month whereas the remaining two (Peaches and Strawberries) are sold only for a few months each year. Figures 1 and 2 plot the prices and quantities of each commodity.

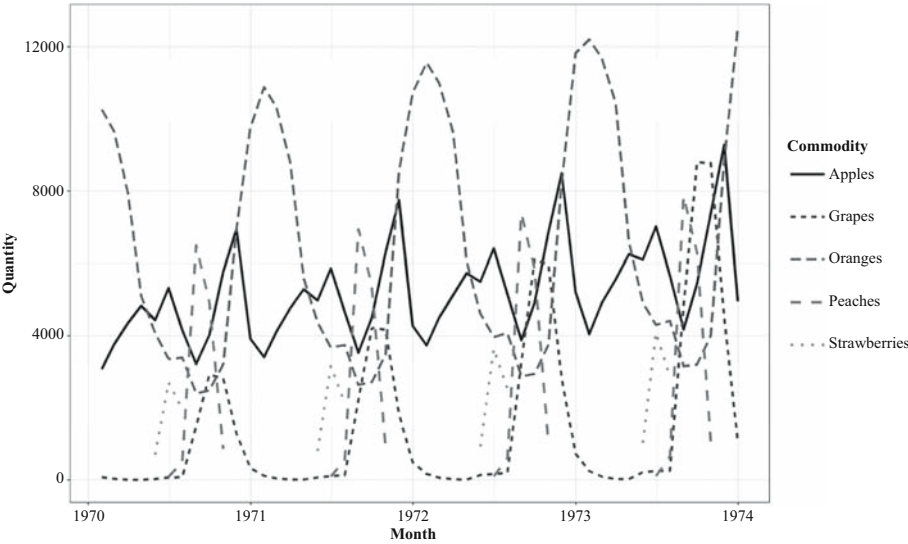


Fig. 2. Quantity of fruit commodities.

Table 2. Features of fruit and scanner data sets.

Commodity class	Fruit	Cookies	Oatmeal	Toothbrushes
Time span used	January 1970 to December 1973	October 1989 to September 1993	July 1991 to June 1995	October 1989 to September 1993
Number of monthly observations	176	18,403	2,617	8,027
Number of commodities	5	763	87	362
Proportion of commodities sold in every month	60%	19%	41%	11%

This Fruit data is useful for our illustration because it contains a small number of products, some of which are not sold every month. However, it does not share all the features of the data to which these methods are applied, such as truly new or disappearing products. For evidence that some of our findings are applicable in practice, we use scanner data for sales of Cookies, Oatmeal and Toothbrushes from Dominick’s stores in Chicago, obtained from the James M. Kilts Center, University of Chicago Booth School of Business. For comparability with the Fruit data, we convert this (weekly) scanner data to monthly frequency by assigning each week to the month in which the majority of its sales fall and subset to 48 months of data; we also remove observations that are flagged as suspect ([University of Chicago 2018](#)). [Table 2](#) summarizes a few features of the Fruit data and the three scanner data sets.

4.2. Multilateral Indexes Calculated on the Entire Fruit Data Set

[Figure 3](#) compares the TPD, CCD and GK price indexes calculated on the entire Fruit data set, with January 1970 as the base period. For this data set, we observe that the three methods produce numerically similar indexes, with the TPD and GK particularly close. Corresponding figures for the other data sets are included as Supplemental Data (Figures A1, A2 and A3). [Figure 4](#) compares the month-on-month price changes corresponding to these indexes.

The multilateral indexes show steep price increases every May. We can use the decomposition methods to understand which commodities are driving these price changes. [Table 3](#) presents the contributions of each commodity to the price change between April and May 1973, using the Simple TPD, Reflexive TPD, CCD and GK decomposition methods presented in this article (recall that the Simple TPD decompositions presented in Subsubsection 2.1.1 are mathematically equivalent). The prices and expenditure shares of each commodity are included for reference. Note that Peaches are not sold in either month, and Strawberries are sold in May but not April.

Overall, the commodity contributions obtained from the Reflexive TPD and GK decompositions are very similar, as would be expected given their mathematical



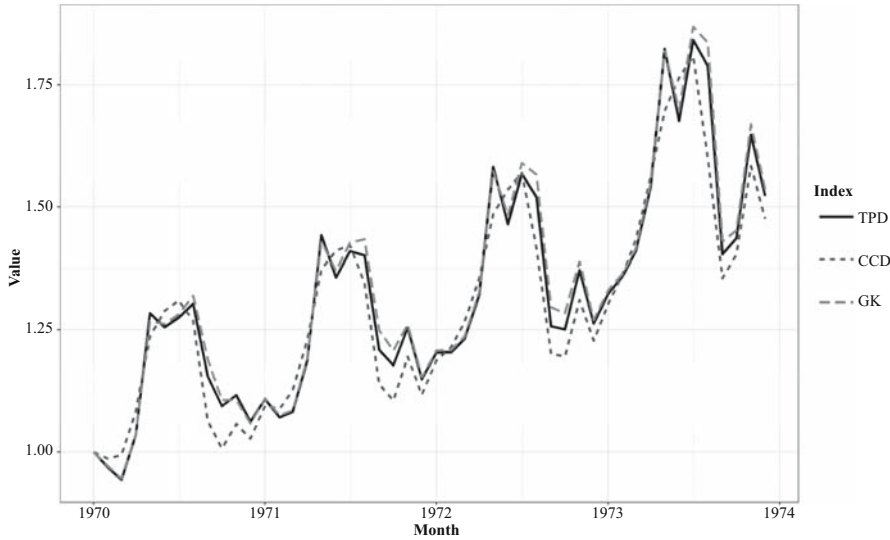


Fig. 3. Multilateral index values.

similarity. It is difficult to draw general conclusions about the numerical similarity of simple and reflexive decompositions of the TPD index, given the latter are not unique—as mentioned in Subsubsection 2.1.2, changing the reference period would yield a different reflexive decomposition.

These contributions reveal a few interesting features of the methods examined. First, the reappearance of Strawberries, a strongly seasonal commodity with an intermittent sales pattern, contributes to an aggregate price increase in May 1973 (it has a contribution greater than one). As this commodity is not sold in the previous month, it

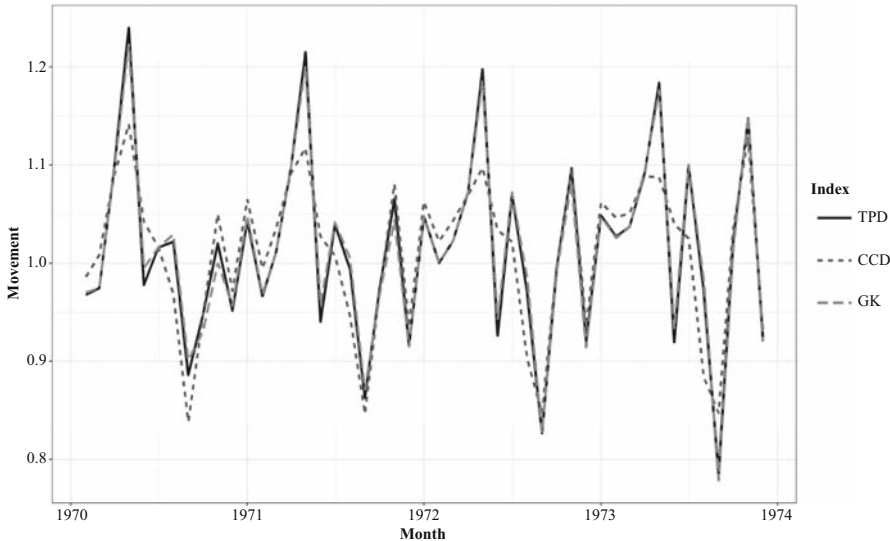


Fig. 4. Month on month multilateral index movements.

Table 3. Decomposition of multilateral index movement between April and May 1973.

Commodity	Contribution to multilateral index movement between April and May 1973				Price		Expenditure share	
	Simple TPD	Reflexive TPD	CCD	GK	April 1973	May 1973	April 1973	May 1973
Apples	1.039	0.993	1.035	1.004	2.00	2.14	0.50	0.42
Grapes	1.017	1.018	0.998	1.016	3.45	3.08	< 0.01	0.02
Oranges	0.997	0.949	1.017	0.956	1.91	2.03	0.50	0.32
Peaches	1.004	1.000	1.000	1.000	NA	NA	NA	NA
Strawberries	1.118	1.233	1.034	1.206	NA	7.17	NA	0.24
Aggregate	1.184	1.184	1.087	1.176			1.00	1.00

would not contribute to a chained bilateral index unless explicit imputation was used. However, the multilateral indexes take into account the prices of this commodity in other periods, compared to which the May 1973 price is relatively high. This capacity to capture the price changes of commodities with intermittent sales is an advantage of using multilateral methods with scanner data: as seen from Table 2, such commodities are common in scanner data sets.

Second, the TPD and GK decompositions show some commodities have price increases between April and May 1973, but contribute to an aggregate price decrease between those periods (contribution less than one) or vice versa. This can occur because the contributions depend on changes in weights, as well as changes in prices. Moreover, the simple TPD decomposition suggests Peaches have a non-trivial contribution to change despite being absent from both periods. These are unintuitive observations, but not disqualifying – by their very definition, multilateral comparisons between two periods take the prices in other periods into account, which may help to mitigate drift (Ivancic et al. 2011) including in the presence of seasonal patterns (Ribe 2012).

Figure 5 shows the relationship between price change and contribution to aggregate price change, for every instance in the Fruit data set where a commodity is sold in consecutive months. They reveal that there is a correlation between commodities' month-on-month price changes and their contribution to change, but also that it is not uncommon for the price changes and contributions to be in opposite directions (observations in the upper left and lower right quadrants). Table 4 illustrates that this phenomenon occurs in scanner data as well. It is consistently less pronounced for the CCD than the TPD index, reflecting that local price changes have greater influence on the CCD index than the TPD index, as observed in Subsection 2.2.

A feature of simple decompositions is that changing the price of one commodity without changing the weights does not affect the contributions of other commodities. Suppose we adjust the price of Oranges in April 1973 to be five times its original value (9.55 instead of 1.91), while leaving the expenditure share unchanged. This price spike might result from adverse production conditions (e.g., a natural disaster), with consumers responding by allocating a fixed expenditure to each commodity and reacting to price changes with reciprocal quantity changes. Observe that this leaves the TPD and CCD weights

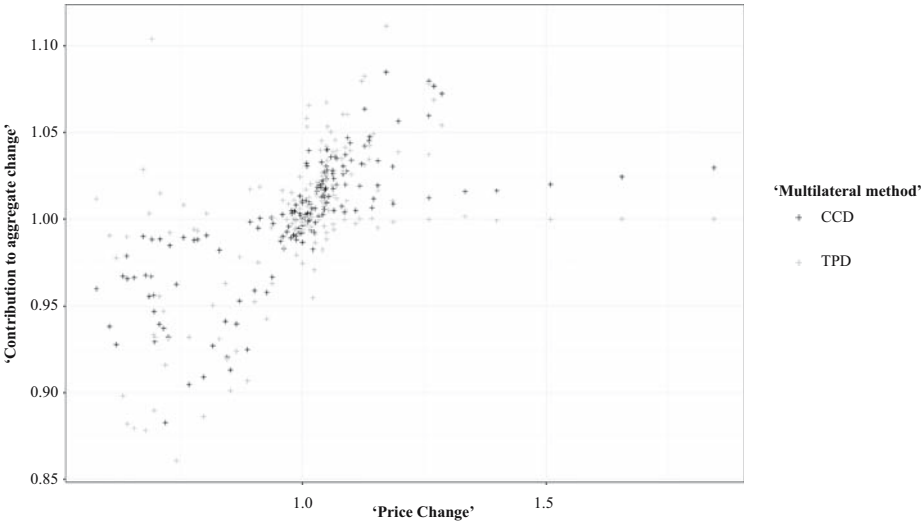


Fig. 5. Commodity price change versus index contribution.

unchanged, but alters the GK weights, so we exclude the latter method from the analysis that follows.

We can recalculate the TPD and CCD indexes using the adjusted data set and derive the contributions of each commodity to the index movement between April and May 1973. Table 4 presents the contributions of each commodity to the index movements from the adjusted data set and their relationship between the commodity contributions from the original data set (in Table 2). For the simple (Simple TPD and CCD) decompositions, only the contribution of Oranges is affected; however, the Reflexive TPD contributions for Apples and Strawberries are slightly altered by the price change of Oranges. This illustrates an advantage of simple decomposition methods.

4.3. Extended Multilateral Indexes

In practice, we would not calculate a multilateral index using the entire data set, but instead use one of the methods described in Section 3 to extend the series one period at a

Table 4. Price changes and contributions in opposite directions.

Commodity class	Fruit	Cookies	Oatmeal	Toothbrushes
Instances where a commodity is sold in consecutive months	165	17,385	2,520	7,467
Instances where contribution is in the opposite direction to price change (Simple TPD)	26%	25%	19%	22%
Instances where contribution is in the opposite direction to price change (CCD)	10%	22%	14%	17%

Table 5. Impact of changing one commodity's price on commodity contributions.

Commodity	Contribution based on adjusted data set			Ratio of adjusted/ original contribution		
	Simple TPD	Reflexive TPD	CCD	Simple TPD	Reflexive TPD	CCD
Apples	1.039	0.990	1.035	1.000	0.997	1.000
Grapes	1.017	1.019	0.998	1.000	1.000	1.000
Oranges	0.448	0.425	0.461	0.449	0.448	0.453
Peaches	1.004	1.000	1.000	1.000	1.000	1.000
Strawberries	1.118	1.239	1.034	1.000	1.005	1.000
Aggregate	0.531	0.531	0.492	0.449	0.449	0.453

time. Figure 6 presents TPD indexes that are extended using the movement splice, window splice, half splice, mean splice, and direct methods. The TPD index based on the entire data set is included for comparison. Corresponding figures for the CCD and GK methods are included as Supplemental Data (Figures A4 and A5). We use a window of length 13 months to calculate the rolling window methods and use a base month of January for the direct method. As the rolling window methods cannot be used to extend the index until a full window of historical data is available, we start the extended indexes in January 1971.

The indexes in Figure 6 are more dispersed than the indexes in Figure 3, indicating that the choice of extension method makes a greater difference to the series than the choice of multilateral method in this example. The mean splice index tracks the index with no extension closely. The direct multilateral index is typically lower than the mean splice in the middle of each calendar year, but similar at the end of the year. The half and movement splice indexes drift a little higher and lower than the mean splice respectively. The window splice index diverges substantially.

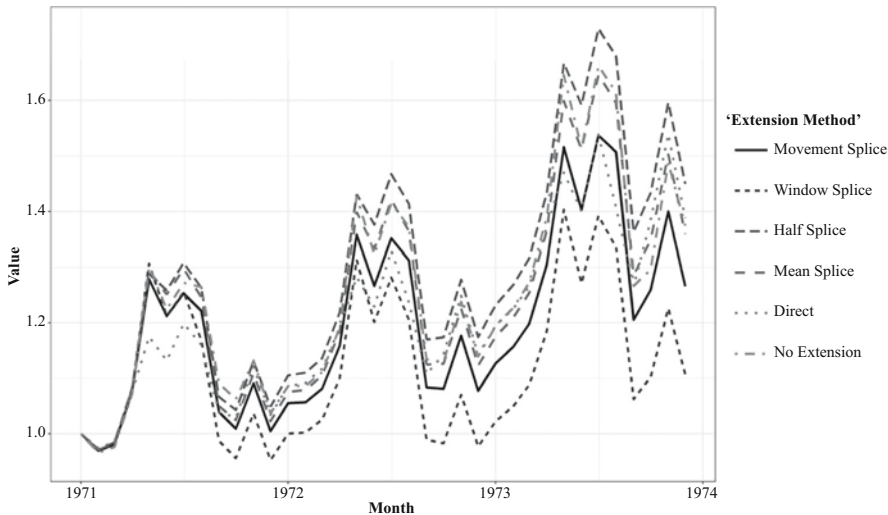


Fig. 6. Extended TPD index values.

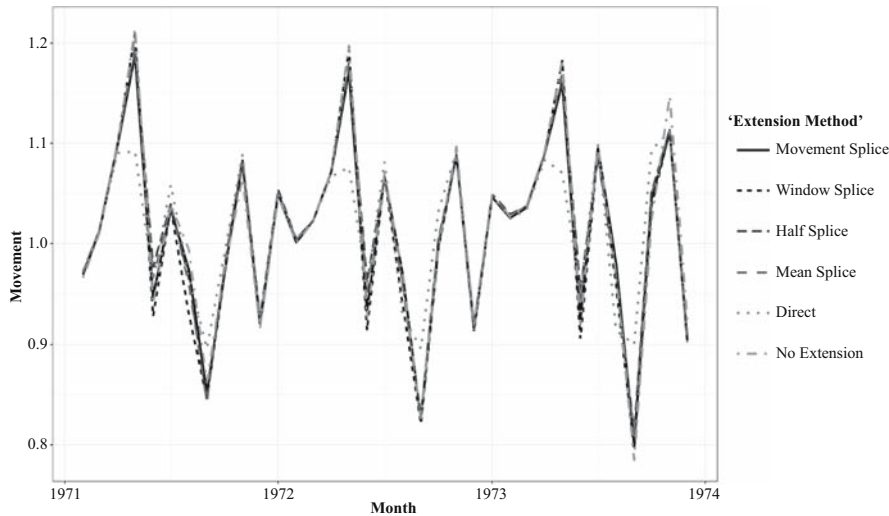


Fig. 7. Month on month extended TPD index movements.

Figures 7 and 8 compare the month-on-month and annual index movements using the various extension methods. The clearest difference in the month-on-month movements is that the direct index movements have a less extreme peak each May and a less extreme trough each September. Otherwise the month-on-month movements appear similar. However, there are systematic differences between the annual movements of the various rolling window methods, implying that these indexes diverge gradually. The decompositions can help to explain these differences.

Table 6 compares the Simple TPD contributions of each commodity to the extended TPD price movements between April and May 1973. The main difference between

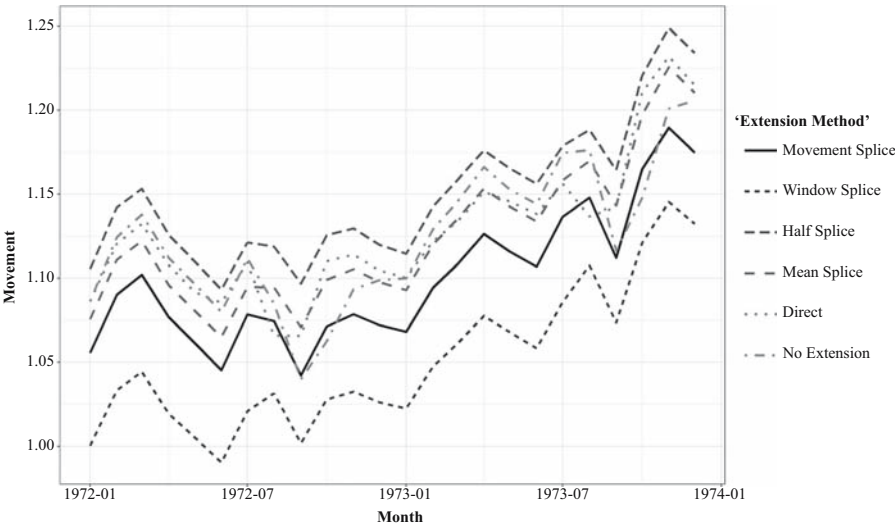


Fig. 8. Annual extended TPD index movements.

Table 6. Decomposition of extended index movement between April and May 1973.

Commodity	Contribution to movement of extended TPD index between April and May 1973					Price		Expenditure share	
	Movement splice	Window splice	Half splice	Mean splice	Direct	April 1973	May 1973	April 1973	May 1973
Apples	1.035	1.034	1.034	1.033	1.034	2.00	2.14	0.50	0.42
Grapes	1.018	1.018	1.020	1.020	1.008	3.45	3.08	<0.01	0.02
Oranges	1.013	1.008	1.013	1.012	1.027	1.91	2.03	0.50	0.32
Peaches	1.003	1.004	1.003	1.004	1.000	NA	NA	NA	NA
Strawberries	1.085	1.109	1.085	1.094	1.000	NA	7.17	NA	0.24
Aggregate	1.161	1.183	1.163	1.170	1.071				

methods is in the contribution of Strawberries. Strawberries do not contribute to the direct index movement between April and May, as the expanding window (starting in January 1973) does not contain any observations for Strawberries until May, and we need two observations for a commodity to contribute to price comparisons.

To understand the differing contributions of Strawberries to the three rolling window methods, note that the high price of Strawberries in May 1973 makes the previous year’s prices in the current window (May 1972 to May 1973) appear lower than they did in the previous window (April 1972 to April 1973). In consequence, the contribution of Strawberries to the price movement between the start of the current window (May 1972) and the previous period (April 1973) is more positive in the current window than in the previous window. As [Krsinich \(2016\)](#) argues, the window splice implicitly revises this price movement in extending the index series from the previous to the current period, whereas the movement splice makes no such revision, and the mean splice makes a partial revision ([Australian Bureau of Statistics 2017](#)). The half splice implicitly revises the movement over part of the previous window (November 1972 to April 1973), but as Strawberries are not sold over this period their contribution to the half splice is the same as their contribution to the movement splice.

[Table 7](#) decomposes the annual movement of each extended TPD index between May 1972 and May 1973. Again, we can see that the commodities with strong seasonality have neutral contributions to the direct index movement because they are not sold between

Table 7. Decomposition of extended index movement between May 1972 and May 1973.

Commodity	Contribution to movement of extended TPD index between May 1972 and May 1973					Price		Expenditure share	
	Movement splice	Window splice	Half splice	Mean splice	Direct	May 1972	May 1973	May 1972	May 1973
Apples	1.026	1.021	1.060	1.040	1.020	1.89	2.14	0.42	0.42
Grapes	0.994	0.984	1.002	0.995	0.999	3.56	3.08	0.02	0.02
Oranges	1.057	1.051	1.086	1.075	1.124	1.70	2.03	0.32	0.32
Peaches	1.010	1.001	1.001	1.009	1.000	NA	NA	NA	NA
Strawberries	1.025	1.011	1.010	1.019	1.000	6.21	7.17	0.24	0.24
Aggregate	1.116	1.068	1.165	1.143	1.146				

January and April of either year. On the other hand, Oranges have a relatively positive contribution to the direct index movement, which likely relates to their high expenditure shares (about 0.75) in January 1972 and January 1973, the base months of the expanding windows used to estimate price changes within those years.

As observed in [Figure 8](#), the half splice has the largest annual movements of the rolling window methods, followed by the mean splice, the movement splice, and the window splice. From [Table 5](#), we can see that the contributions of the commodities that are sold all year round (Apples, Grapes and Oranges) follow the same ordering. Note that Peaches make a non-trivial contribution to the rolling window index movements — even though they are not sold in May, their prices in intervening months contribute to the month-on-month movements of the extended index, and ultimately to the annual movement.

## 5. Conclusions

Index decomposition is useful in practice for interpreting price indexes: it allows one to break down aggregate price movements into contributions from individual or groups of commodities. Decomposition is particularly important for understanding multilateral indexes, which combine many different time comparisons yielding complex dependencies on any individual commodity's price observations. We defined reflexive or simple decompositions based on whether the contribution for each commodity depends on an aggregate price level. Simple decompositions ensure contributions for commodities are invariant under changes in other commodity prices.

We introduced a simple decomposition for the TPD index and a reflexive decomposition for the GK index, and showed how these compare to the reflexive decomposition for the TPD index and the simple decomposition for the CCD index. These decompositions demonstrate that movements can be attributed to the price observations for each commodity. The theoretical and empirical results provide evidence of similarities between these three indexes and subtle differences between the CCD and the other methods. They also show how commodities sold in only one of two time periods can influence the price comparison between those periods, and reveal that is not uncommon for a commodity's contribution to aggregate price change to be in the opposite direction to its individual price change.

The comparison between decompositions raises questions for price index implementations. Where decompositions disagree on the direction of the effect of particular commodities, how should this be interpreted? Under what conditions should commodity contributions remain invariant?

We do not fully address these questions here. Where several decomposition methods are available, each may yield additional information about price movements. The circumstances in which the price index is applied may dictate which decomposition is most useful, such as the choice between an additive and a multiplicative method.

However, we have touched on several properties that it seems advantageous for a decomposition to possess, including that one commodity's contribution should be invariant to changes in the prices of other commodities (conditional on the expenditure shares), and that a commodity not sold in either of two periods should have a trivial contribution to the price change between those periods. Other desirable properties might

include invariance to the ordering of commodities and time periods, or invariance to the price changes of other commodities under different conditions. Development of a more complete set of desirable properties for index decomposition functions would be an interesting area for further research.

## 6. References

- Australian Bureau of Statistics (ABS). 2016. "Making Greater Use of Transactions Data to Compile the Consumer Price Index, Australia." Cat. no. 6401.0.60.003. Canberra: ABS. Available at: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/mf/6401.0.60.003> (accessed May 2019).
- Australian Bureau of Statistics (ABS). 2017. "An Implementation Plan to Maximise the Use of Transactions Data in the CPI." Cat. no. 6401.0.60.004. Canberra: ABS. Available at: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/mf/6401.0.60.004> (accessed May 2019).
- Balk, B.M. 2008. *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*. New York: Cambridge University Press. Doi: <https://doi.org/10.1017/cbo9780511720758>.
- Caves, D.W., L.R. Christensen, and W.E. Diewert. 1982. "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers." *The Economic Journal* 92(365): 73–86. Doi: <http://dx.doi.org/10.2307/2232257>.
- Chessa, A. 2015. "Towards a Generic Price Index Method for Scanner Data in the Dutch CPI." Paper presented at the 14th meeting of the Ottawa Group, Tokyo, 20–22 May 2015. Available at: <http://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s1room2.pdf> (accessed August 2017).
- Chessa, A., J. Verburg, and L. Willenborg. 2017. "A Comparison of Price Index Methods for Scanner Data." Paper presented at the 15th meeting of the Ottawa Group, Eltville, 10–12 May 2017. Available at: [https://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research\\_Centre/Conferences/2017/2017\\_05\\_10\\_ottawa\\_group\\_07\\_1\\_paper.pdf](https://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research_Centre/Conferences/2017/2017_05_10_ottawa_group_07_1_paper.pdf) (accessed February 2018).
- Collier, I.L. 1999. "Comment." *International and Interarea Comparisons of Income, Output, and Prices*, edited by A. Heston and R. E. Lipsey, 87–107. Chicago: University of Chicago Press. Doi: <https://doi.org/10.7208/chicago/9780226331126.001.0001>.
- De Haan, J. 2015. "A Framework for Large Scale Use of Scanner Data in the Dutch CPI." Paper presented at the 14th meeting of the Ottawa Group, Tokyo, 20–22 May 2015. Available at: [http://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s1lp33\\_paper.pdf](http://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s1lp33_paper.pdf) (accessed August 2017).
- De Haan, J., R. Hendriks, and M. Scholz. 2016. "A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes." Paper presented at the 15th meeting of the Ottawa Group, Eltville, 10–12 May 2017. Available at: <https://www.bundesbank.de/resource/blob/636054/f20b4679b121998a5ae5f106a11ee5a1/mL/2017-05-10-ottawa-group-07-4-paper-data.pdf> (accessed February 2019).
- De Haan, J. and F. Krsinich. 2014. "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes." *Journal of Business & Economic Statistics* 32(3): 341–358. Doi: <http://dx.doi.org/10.1080/07350015.2014.880059>.



- Diewert, W.E. 2002. "The Quadratic Approximation Lemma and Decompositions of Superlative Indexes." *Journal of Economic and Social Measurement* 28(1,2): 63–88. Doi: <https://doi.org/10.3233/JEM-2003-0200>.
- Diewert, W.E. and K.J. Fox. 2017. "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data." Paper presented at the 15th meeting of the Ottawa Group, Eltville, 10–12 May 2017. Available at: <https://www.bundesbank.de/resource/blob/635970/864de61dee8f2b67dcb3ed2a2ab479c5/mL/2017-05-10-ottawa-group-07-2-paper-data.pdf> (accessed February 2019).
- Éltető, Ö., and P. Köves. 1964. "On an Index Number Computation Problem in International Comparison" (in Hungarian). In *Statistikai Szemle*, 42: 507–18. Available at: [https://www.ksh.hu/statszemle\\_archivum#year=1964/issue=05](https://www.ksh.hu/statszemle_archivum#year=1964/issue=05) (accessed May 2019).
- Geary, R.C. 1958. "A Note on the Comparison of Exchange Rates and Purchasing Power Between Countries." *Journal of the Royal Statistical Society. Series A (General)* 121(1): 97–99. Doi: <http://dx.doi.org/10.2307/2342991>.
- Gini, C. 1931. "On the Circular Test of Index Numbers." *Metron* 9(9): 3–24.
- Howard, A., K. Dunford, J. Jones, M. van Kints, K. Naylor, and R. Tarnow-Mordi. 2015. "Using Transactions Data to Enhance the Australian CPI." Paper presented at the 14th meeting of the Ottawa Group, Tokyo, 20–22 May 2015. Available at: [http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/d012f001b8a1cf6cca257eed008074c9/\\$FILE/Australian\\_Bureau\\_of\\_Statistics-Using\\_transactions\\_data\\_to\\_enhance\\_the\\_Australian\\_CPI.pdf](http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/d012f001b8a1cf6cca257eed008074c9/$FILE/Australian_Bureau_of_Statistics-Using_transactions_data_to_enhance_the_Australian_CPI.pdf) (accessed August 2017).
- ILO, IMF, OECD, UNECE, Eurostat, The World Bank. 2004. *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Organization. Doi: <https://doi.org/10.5089/9789221136996.069>.
- Ivancic, L., W. Erwin Diewert, and K.J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes." *Journal of Econometrics* 161(1): 24–35. Doi: <http://dx.doi.org/10.1016/j.jeconom.2010.09.003>.
- Khamis, S.H. 1972. "A New System of Index Numbers for National and International Purposes." *Journal of the Royal Statistical Society. Series A (General)* 135(1): 96–121. Doi: <http://dx.doi.org/10.2307/2345041>.
- Krsinich, F. 2016. "The FEWS Index: Fixed Effects with a Window Splice." *Journal of Official Statistics* 32(2): 375. Doi: <http://dx.doi.org/10.1515/jos-2016-0021>.
- Punatanen, S., and G.P.H. Styan. 2006. "Historical Introduction: Issai Schur and the Early Development of the Schur Complement." In *The Schur Complement and its Applications*, edited by F. Zhang. 1–16. New York: Springer. Doi: [https://doi.org/10.1007/0-387-24273-2\\_1](https://doi.org/10.1007/0-387-24273-2_1).
- Rao, D.S.P. 1990. "A System of Log-Change Index Numbers for Multilateral Comparisons." In *Contributions to Economic Analysis* 194: 127–139. Doi: <https://doi.org/10.1016/b978-0-444-88409-1.50013-0>.
- Rao, D.S.P. 2005. "On the Equivalence of Weighted Country-Product-Dummy (CPD) Method and the Rao-System For Multilateral Price Comparisons." *Review of Income and Wealth* 51(4): 571–80. Doi: <http://dx.doi.org/10.1111/j.1475-4991.2005.00169.x>.

- Reinsdorf, M.B., W.E. Diewert, and C. Ehemann. 2002. "Additive Decompositions for Fisher, Törnqvist and Geometric Mean Indexes." *Journal of Economic and Social Measurement* 28(1,2): 51–61. Doi: <https://doi.org/10.3233/jem-2003-0194>.
- Ribe, M. 2012. "Some Properties of the RGEKS Index for Scanner Data." Paper presented at Statistics Sweden's scanner data workshop, 7–8 June 2012, Stockholm, Sweden. Available at: [https://www.scb.se/Statistik/PR/PR0101/\\_dokument/Some%20properties%20of%20the%20RGEKS%20index%20for%20scanner%20data.pdf](https://www.scb.se/Statistik/PR/PR0101/_dokument/Some%20properties%20of%20the%20RGEKS%20index%20for%20scanner%20data.pdf) (accessed February 2019).
- Summers, R. 1973. "International Price Comparisons based upon Incomplete Data." *Review of Income and Wealth* 19(1): 1–16. Doi: <http://dx.doi.org/10.1111/j.1475-4991.1973.tb00870.x>.
- Szulc, B. 1964. "Index Numbers of Multilateral Regional Comparisons" (in Polish). In *Przegląd Statystyczny*, 3: 239–54.
- Turvey, R. 1979. "Treatment of Seasonal Items in Consumer Price Indices." In *Bulletin of Labor Statistics*, 4, 13–23. Geneva: ILO. Available at: [https://www.ilo.org/public/libdoc/ilo/P/09606/09606\(1979-4\)XIII-XXIII.pdf](https://www.ilo.org/public/libdoc/ilo/P/09606/09606(1979-4)XIII-XXIII.pdf) (accessed May 2019).
- University of Chicago. 2018. "Kilts Center for Marketing: Dominick's Data Manual." Chicago: University of Chicago Booth School of Business. Available at: [https://www.chicagobooth.edu/-/media/enterprise/centers/kilts/datasets/dominicks-dataset/dominicks-manual-and-codebook\\_kiltscenter.aspx](https://www.chicagobooth.edu/-/media/enterprise/centers/kilts/datasets/dominicks-dataset/dominicks-manual-and-codebook_kiltscenter.aspx) (accessed November 2018).
- Van IJzeren, J. 1952. "On the Plausibility of Fisher's Ideal Indices" (in Dutch). In *Statistische en Econometrische Onderzoekingen, Nieuwe Reeks*, 7, 104–15. The Hague: CBS.
- Van IJzeren, J. 1983. "Index Numbers for Binary and Multilateral Comparison: Algebraical and Numerical Aspects." In *Statistical Studies*, 34. The Hague: CBS.
- Vartia, Y.O. 1974. "Relative Changes and Economic Indices." Doctoral dissertation, University of Helsinki.
- Vartia, Y.O. 1976. "Ideal Log-Change Index Numbers." *Scandinavian Journal of Statistics* 3(3): 121–126. Available at: <http://www.jstor.org/stable/4615624> (accessed May 2019).

Received June 2018

Revised December 2018

Accepted February 2019