ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

The Curious Case of an Exploding GEKS Index

Kevin J. Fox

ESCoE Discussion Paper 2026-12

August 2026

ISSN 2515-4664

DISCUSSION PAPER

The Curious Case of an Exploding GEKS Index
Kevin J. Fox
ESCoE Discussion Paper No. 2026-12
August 2026

Abstract

Multilateral indexes are increasingly being used to incorporate transactions data into consumer price indices. This paper shows how the multilateral GEKS-Fisher index can misbehave in a spectacular fashion when using such data. It is found that this is due to observations with very high price relatives. In contrast, the GEKS-Törnqvist index is very resilient to extreme observations. It is also found that the GEKS-Fisher index can be resilient, if implemented using an appropriately sized rolling window. The choice of splicing method, used for updating as new data becomes available, is found to be immaterial in this respect.

Keywords: price indexes, multilateral index numbers, scanner data, chain drift, su-perlative indexes

JEL classification: C43, C82, E31

Kevin J. Fox, UNSW Sydney and ESCoE
k.fox@unsw.edu.au.

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

The Curious Case of an Exploding GEKS Index

Kevin J. Fox*

UNSW Sydney and ESCoE

August 2026

Abstract

Multilateral indexes are increasingly being used to incorporate transactions data into consumer price indices. This paper shows how the multilateral GEKS-Fisher index can misbehave in a spectacular fashion when using such data. It is found that this is due to observations with very high price relatives. In contrast, the GEKS-Törnqvist index is very resilient to extreme observations. It is also found that the GEKS-Fisher index can be resilient, if implemented using an appropriately sized rolling window. The choice of splicing method, used for updating as new data becomes available, is found to be immaterial in this respect.

Keywords: price indexes, multilateral index numbers, scanner data, chain drift, superlative indexes

JEL classification: C43, C82, E31

*School of Economics and Centre for Applied Economic Research, UNSW Sydney. Email: K.Fox@unsw.edu.au. I thank Erwin Diewert, Jan de Haan, Peter Levell, Daniel Melser and Martin O'Connell for helpful discussions and comments, Graham White for making the `IndexNumR` package publicly available, and participants at the UNSW-ESCoE Economic Measurement Conference (Sydney, December 2025) and at the 19th Ottawa Group meeting (Warsaw, May 2026) for helpful comments. All errors are my own.

1 Introduction

National statistical agencies are increasingly using transactions data, such as supermarket scanner data, to compile consumer price indices (CPIs). Such data record the price and quantity of every item sold in every period. Unlike traditional CPI price surveys, scanner data typically exhibit substantial product churn, as items enter and exit the marketplace frequently. Prices, quantities and expenditure shares can move sharply from one period to the next. It has been shown that when using such data, chained bilateral indexes can generate substantial *chain drift bias*, even when using "superlative" indexes such as Fisher or Törnqvist (de Haan, 2008; Ivancic et al., 2011; de Haan and van der Grient, 2011).

Multilateral index number methods, and in particular the GEKS index (Gini, 1931, Eltëtö and Köves, 1964 and Szulc, 1964), were recommended as a solution to chain drift bias in this setting by Ivancic et al. (2011), and have subsequently been adopted for official CPI compilation by a number of national statistical agencies, including those of Australia, Austria, Belgium, Canada, Luxembourg, the Netherlands, New Zealand, Norway, and the United Kingdom. A GEKS index between any two periods is constructed as the geometric mean, over all periods in an estimation window, of the indirect comparisons between the two periods obtained by using each period in the window in turn as the base period for bilateral comparisons. When the bilateral index used is the Fisher ideal index, the result is referred to as the GEKS-Fisher index; when the Törnqvist index is used, the result is referred to as the GEKS-Törnqvist index, also known as the CCDI index after Caves et al. (1982) and Inklaar and Diewert (2016).[1]

Standard superlative indexes closely approximate one another numerically, typically agreeing to the second or third decimal place (Diewert, 1978; Hill, 2006). It might therefore be expected that GEKS-Fisher and GEKS-Törnqvist indexes, being built from these closely-approximating bilateral building blocks, would themselves closely approximate one another empirically. This paper provides a striking counterexample using real-world scanner data: the GEKS-Fisher index *explodes*, producing an implausible spike of more than twenty percentage points in a single month, while the GEKS-Törnqvist (CCDI) index shows no such behaviour.

This finding complements two recent contributions. Fox et al. (2026, p. 391) noted the empirical instability of GEKS-Fisher indexes when using high-frequency data, in the context of measuring inflation during a period of rapidly changing relative prices. Dikhanov (2025) examined, from a theoretical perspective, the potential divergence between Fisher

[1] Caves et al. (1982) set out its properties from the economic approach to index numbers in the context of quantity and productivity indexes, while Inklaar and Diewert (2016) extended the economic framework to price indexes.

and Törnqvist indexes when relative price changes are large. This paper provides a fully worked empirical example using publicly available scanner data, rather than a simulation. It traces the mechanism responsible for the divergence down to the level of an individual transaction. It evaluates three practical remedies: (i) using the GEKS-Törnqvist (CCDI) index in place of GEKS-Fisher; (ii) implementing GEKS-Fisher using an appropriately sized rolling window; and (iii) data cleaning to remove the offending observation(s).

2 Bilateral and Multilateral Index Number Methods

Let p^0 and p^1 denote vectors of prices, and q^0 and q^1 vectors of quantities, for N products in periods 0 and 1 respectively, and let $s_n^t \equiv p_n^t q_n^t/(p^t \cdot q^t)$ denote the expenditure share of product n in period t. The Laspeyres (P_L) and Paasche (P_P) bilateral price indexes are as follows:

$$P_L(p^0, p^1, q^0, q^1) \equiv \frac{p^1 \cdot q^0}{p^0 \cdot q^0} = \sum_{n=1}^{N} s_n^0 \left(\frac{p_n^1}{p_n^0}\right), \tag{1}$$

$$P_P(p^0, p^1, q^0, q^1) \equiv \frac{p^1 \cdot q^1}{p^0 \cdot q^1} = \left[\sum_{n=1}^{N} s_n^1 \left(\frac{p_n^1}{p_n^0}\right)^{-1}\right]^{-1}. \tag{2}$$

The Fisher ideal index (P_F) is the geometric mean of the Laspeyres and Paasche indexes,

$$P_F(p^0, p^1, q^0, q^1) \equiv [P_L\, P_P]^{1/2}, \tag{3}$$

and the Törnqvist index (P_T) is a share-weighted geometric mean of the price relatives, where the weight on each product is the average of its expenditure share in the two periods,

$$\ln P_T(p^0, p^1, q^0, q^1) = \sum_{n=1}^{N} \tfrac{1}{2}\left(s_n^0 + s_n^1\right) \ln\left(\frac{p_n^1}{p_n^0}\right). \tag{4}$$

Both P_F and P_T are *superlative* indexes in the sense of Diewert (1976), in that each is exact for a flexible functional form.[2]

The GEKS multilateral index, originally developed for making transitive comparisons across countries, can equally be applied in a time-series context over a window of $T+1$ periods, $t = 0, \ldots, T$. Transitivity means that the direct comparison between any two periods equals the comparison obtained by chaining through any third period. Let $P_{b,t}$

[2] A functional form is "flexible" if it can provide a second order approximation to an arbitrary twice differentiable linearly homogeneous function at a point (Diewert, 1974, p. 113).

denote a bilateral price index for period t using period b as the base period; for example, $P_{b,t} = P_F(p^b, p^t, q^b, q^t)$ for the Fisher index. The GEKS index for period t, relative to period 0, is then defined as follows:

$$GEKS_{0,t} = \prod_{b=0}^{T} \left[\frac{P_{b,t}}{P_{b,0}} \right]^{1/(T+1)}. \tag{5}$$

The interpretation is as follows. Each period b in the window is used, in turn, as the base period for a pair of bilateral indexes: $P_{b,t}$ compares period t with base period b, and $P_{b,0}$ compares period 0 with the same base period. Their ratio, $P_{b,t}/P_{b,0}$, is an indirect comparison of period t with period 0, made through base period b. The GEKS index is the geometric mean of the $T + 1$ indirect comparisons obtained by using every period in the window as the base period. It is this symmetric treatment of all possible base periods that frees the index from dependence on any single base period and delivers transitivity. If the bilateral index satisfies the time reversal test, $P_{b,0} = 1/P_{0,b}$, as both the Fisher and Törnqvist indexes do, then equation (5) can be written equivalently as $GEKS_{0,t} = \prod_{b=0}^{T} [P_{0,b} \times P_{b,t}]^{1/(T+1)}$, a geometric mean of chained comparisons passing from period 0 to period b, and then from period b to period t. When the underlying bilateral index in equation (5) is Fisher, this yields the GEKS-Fisher index, and when it is Törnqvist this yields the GEKS-Törnqvist (CCDI) index.

A GEKS calculation over the whole sample revises the entire index every time a new observation becomes available. This is unacceptable for a CPI that must be non-revisable. The now standard solution is to compute the GEKS index over a rolling window of fixed length (for example the most recent 13 or 25 months), and to splice each new period onto the previous index using only the movement implied by the new window. Several splicing methods have been proposed and compared in the literature.[3]

A benchmark GEKS index is also computed using (5) with a single window spanning the entire sample period. This index is a useful analytical device for two reasons. First, as no updating is done, any anomalous behaviour cannot be due to the use of a particular updating method. Second, it gives the index maximum exposure to problematic comparisons, as each period is compared with every other period of the sample, with each period used in turn as the base period b in equation (5).

[3] See e.g. Ivancic et al. (2011), de Haan (2015), Krsinich (2016), Melser (2018), Diewert and Fox (2022) and Fox et al. (2026). This paper uses the mean splice proposed by Ivancic et al. (2011, p. 33, footnote 19) and explored by Diewert and Fox (2022), in which the geometric mean of the period-on-period movements implied by all of the overlapping periods common to the old and new windows is used.

3 Data

The data are from the Dominick's Finer Foods database, covering supermarket transactions in the Chicago metropolitan area, made publicly available by the Kilts Center for Marketing at the University of Chicago Booth School of Business. Data for grooming products are used. This is one of a number of product categories available in the database, and is used here because it displays the anomalous behaviour that is the subject of this paper. The aim is to establish that such behaviour can arise with real transactions data, and to identify its cause. Following standard statistical agency practice (ILO et al., 2020), weekly transaction records are aggregated to form average monthly prices by taking unit values (expenditure divided by quantity sold) for each product, where products are defined by their Universal Product Code (UPC).[4]

The resulting monthly panel spans 64 months, with 1,381 unique products and 32,139 product-month transactions in total. To get an idea of the nature of the data, three figures are presented. As is typical of scanner data, the product set exhibits substantial churn. Figure 1 shows the number of newly appearing and newly disappearing products in each month, both of which fluctuate around 20–40 in most months but spike sharply above 150 in several months. This is most striking around months 29, 35, 50 and 53–54, and from Figure 2 we see that these coincide with large swings in the total number of products transacted per month. Figure 3 shows the expenditure share in each month that is accounted for by products that also existed in the first period, which falls from close to 90 per cent in the early months to under 30 per cent by the end of the sample. Such product churn causes problems for index numbers, as they require a comparison of at least two periods. It is problematic for standard matched-model bilateral indexes, where each comparison requires the same product list in the two periods being compared.[5] Hence, it is also problematic for multilateral indexes composed from such bilateral indexes (Diewert et al., 2025). Product churn can cause chain drift bias in chained bilateral indexes, and re-introduce chain drift bias into multilateral indexes updated with a rolling window, even though each multilateral index window is free from chain drift; see e.g. Fox et al. (2026).

While these figures reveal that there is significant product churn that can cause problems in constructing reliable index numbers, perhaps contrary to prior expectations, we will see in Section 4 that standard plots such as these do not necessarily help in identifying the problem which will cause the GEKS-Fisher index to "explode".

[4]The R package `IndexNumR` (White, 2021) provides the underlying implementations of the index number methods used in this paper.

[5]Fixed-base bilateral index comparisons are particularly affected by product churn, becoming increasingly unreliable as time passes and the number of matched models approaches zero.

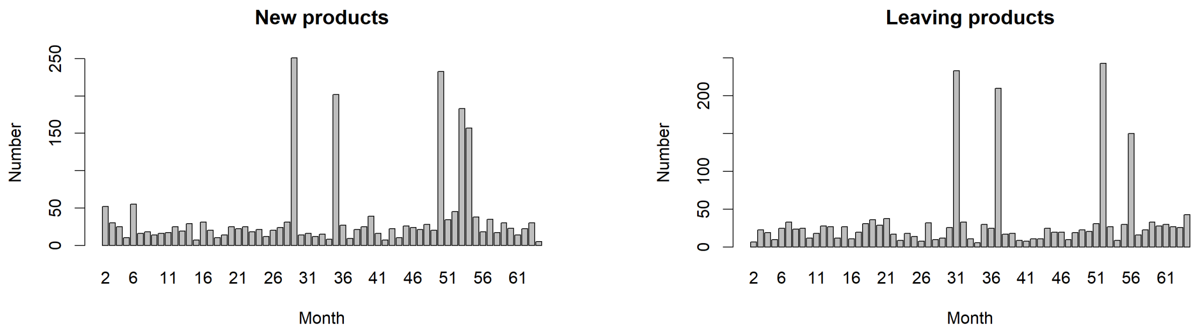

Figure 1: Number of new (left) and leaving (right) products by month, grooming products category.

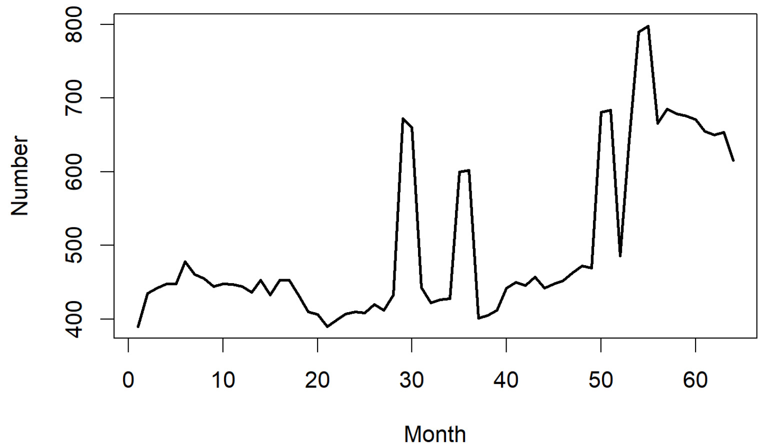

Figure 2: Total number of products transacted by month, grooming products category.

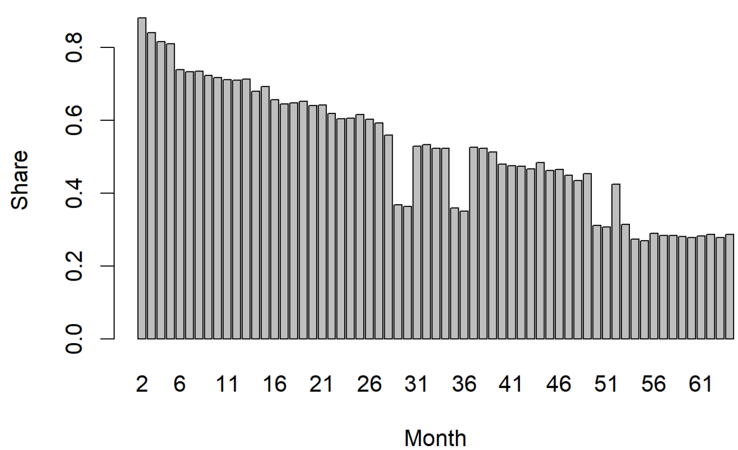

Figure 3: Expenditure share of products in each month that also transacted in the base (first) period, grooming products category.

4 An Exploding GEKS-Fisher Index

Figure 4 plots fixed-base Laspeyres, Paasche, Fisher and Törnqvist indexes for the grooming products category, normalised to 1 in the first month. As is typically the case (due to substitution bias), the Laspeyres index lies above the Paasche index throughout, while the Fisher and Törnqvist indexes lie between the two and track one another almost exactly, to the extent that they are barely distinguishable in the figure. Figure 5 shows the corresponding chained versions of these same indexes. The chained Laspeyres index drifts sharply upward and the chained Paasche index drifts sharply downward relative to the fixed-base indexes in Figure 4, providing a clear example of chain drift bias. The chained Fisher and chained Törnqvist indexes are comparatively well-behaved and continue to track one another closely, consistent with the typical numerical closeness of these superlative bilateral indexes.

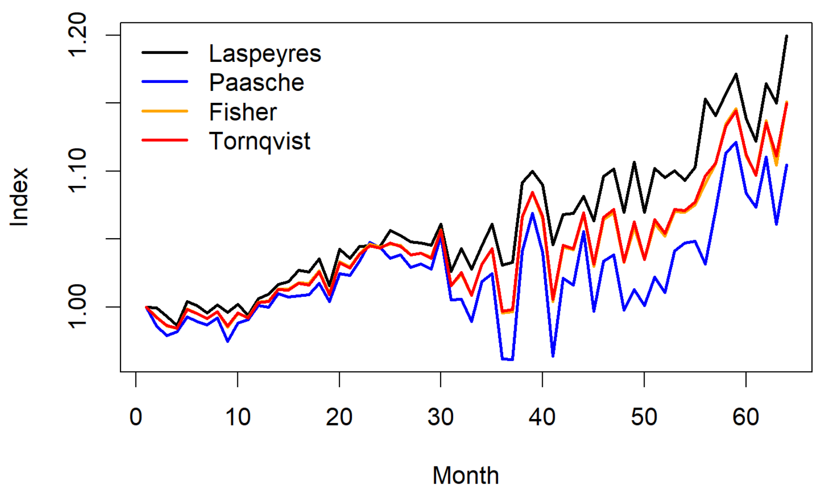

Figure 4: Fixed-base Laspeyres, Paasche, Fisher and Törnqvist indexes

Figure 6 compares the (full-sample benchmark) multilateral GEKS-Törnqvist (CCDI) index with the bilateral fixed-base Törnqvist index from Figure 4. Both are free from chain drift. The two track one another closely. This is reassuring in that the CCDI index, despite using a different method of construction, gives results very similar to the simple fixed-base Törnqvist index, and shows no sign of the kind of erratic behaviour associated with chain drift.

The GEKS-Fisher index is dramatically different. Figure 7 plots the benchmark CCDI index alongside the benchmark GEKS-Fisher index, together with mean-spliced, rolling-window versions of each (to be discussed later in this section). While the CCDI benchmark index is smooth throughout, the GEKS-Fisher benchmark index exhibits an enormous, iso-

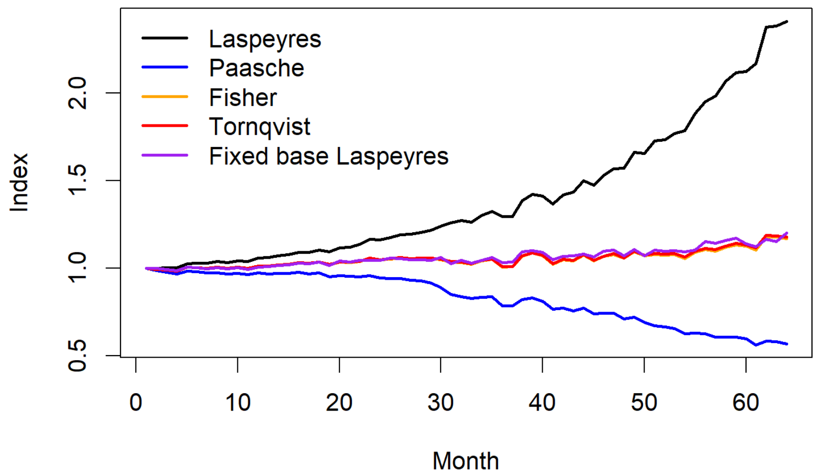

Figure 5: Chained Laspeyres, Paasche, Fisher and Törnqvist indexes, together with the fixed-base Laspeyres index. The chained Laspeyres and Paasche indexes illustrate substantial chain drift.

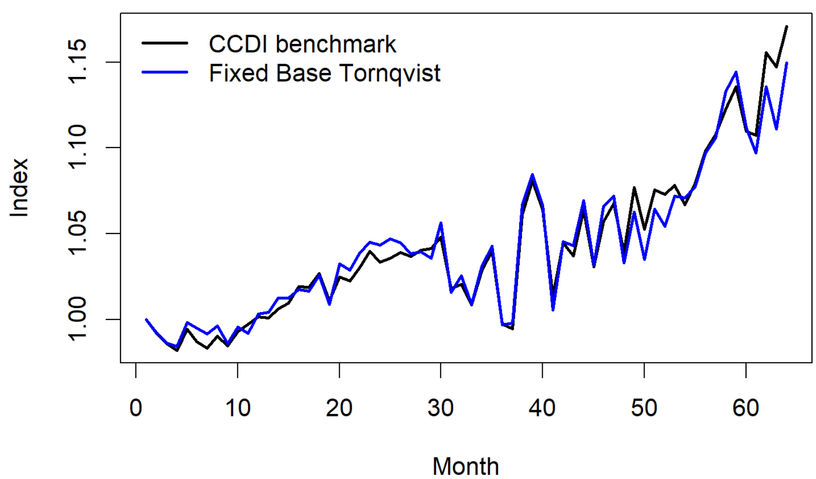

Figure 6: Benchmark GEKS-Törnqvist (CCDI) index compared with the fixed-base Törnqvist index.

lated spike at month 11, reaching a level of approximately 1.22. This is more than twenty percentage points above the level implied by every other index in the figure, and by the surrounding months of the GEKS-Fisher series itself. The spike appears and disappears within a single month. No other month in the 64-month sample shows any comparable movement.

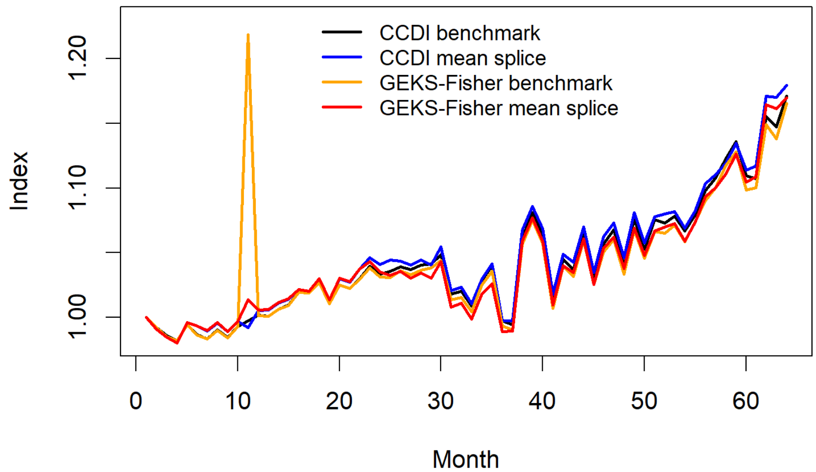

Figure 7: Benchmark and mean-spliced CCDI (GEKS-Törnqvist) and GEKS-Fisher indexes. The GEKS-Fisher benchmark index (full-sample window) spikes sharply at month 11.

To find what is causing this behaviour, Figure 8 decomposes the benchmark GEKS-Fisher index into its GEKS-Laspeyres and GEKS-Paasche building blocks. These are constructed by applying the GEKS geometric-mean-of-bilateral-comparisons formula in equation (5) to the Laspeyres and Paasche indexes in equations (1) and (2), respectively.[6] The GEKS-Laspeyres index spikes to almost 1.49 at month 11, an implausible 49 per cent increase in the price of grooming products in a single month. The GEKS-Paasche index remains smooth and unremarkable throughout, at close to 1.0 in month 11. Because the GEKS-Fisher index is the geometric mean of these two component indexes, it is pulled substantially upward by the extreme GEKS-Laspeyres value.

The source of the spike can be traced to a single product, one of the 447 transacted in month 11: UPC 4740000000, Right Guard Gel Antiperspirant, 3oz. This item first appears in month 11, when a single unit is sold for \$403.50. This implies a unit value roughly 60 to 200 times its subsequent levels (see Figure 10 below). It almost certainly reflects either a data error, or a one-off transaction; for example, an incorrectly labelled sale of multiple

[6]The `IndexNumR` implementation of the GEKS algorithm exploits the time reversal property, $P_{b,t} = 1/P_{t,b}$ for any periods b and t, of the Fisher and Törnqvist indexes to halve the number of bilateral comparisons that must be computed. As Laspeyres and Paasche indexes do not satisfy time reversal, the full $(T+1) \times (T+1)$ matrix of bilateral comparisons was computed directly for this decomposition exercise.

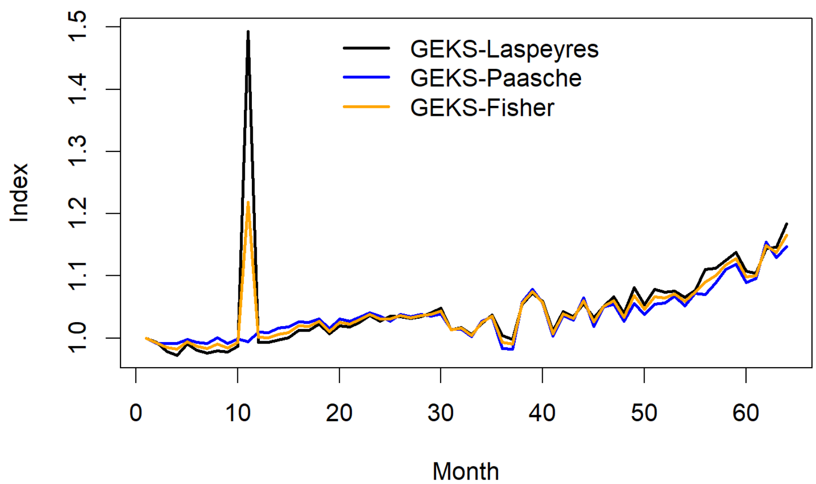

Figure 8: Benchmark GEKS-Laspeyres, GEKS-Paasche and GEKS-Fisher indexes. The GEKS-Laspeyres index spikes to nearly 1.49 at month 11, driving the corresponding spike in GEKS-Fisher.

units in a bundle. The item then disappears until month 22, after which it reappears with a much lower unit value. Specifically, 7 units are sold at a unit value of $2.86. From this point on the unit value of the item remains fairly stable at this level, but the quantity sold grows rapidly, reaching 186 units by month 23 and continuing to grow thereafter. From Figure 9, it can be seen that the item's share of expenditure grows from zero to as much as 0.64 per cent of the entire grooming products category by the second half of the sample.

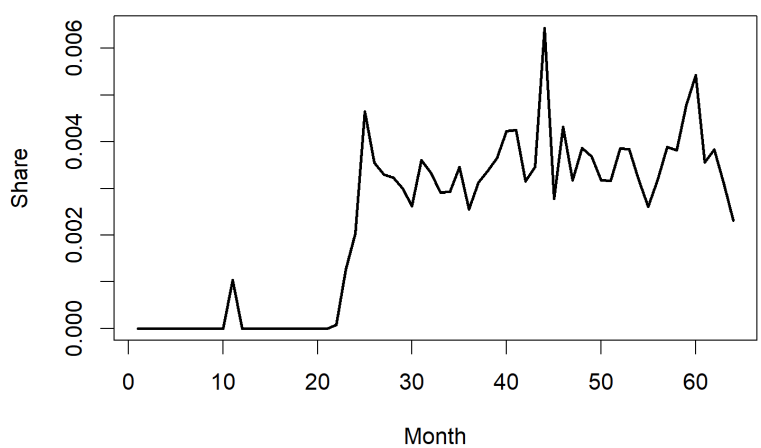

Figure 9: Expenditure share of UPC 4740000000.

This combination of a single extreme unit value in month 11 followed by a rapidly growing expenditure share in later months causes the spike in the Laspeyres index, and hence in the GEKS-Fisher. Recall from equation (1) that the Laspeyres index uses base-period expenditure shares as the weights on the price relatives. For UPC 4740000000, the contribution to the Laspeyres index $P_{b,11}$, comparing month 11 with base period b, is

$$(\text{month-}b\text{ share}) \times \left(\frac{\text{month-11 unit value}}{\text{month-}b\text{ unit value}}\right). \tag{6}$$

Because a full-sample benchmark GEKS index uses every one of the 64 months as a base period in turn (equation (5)), this comparison is made 64 times, including many times against months from month 22 onward, by which point the item has a non-trivial and growing expenditure share. Figure 10 shows, for each possible base period, both the raw price relative (panel a) and the base-period-share-weighted price relative (panel b) for month 11 relative to that base. The raw price relative for this item ranges from around 60 to nearly 200 (i.e., the month-11 unit value is up to 200 times the unit value in the given base period). Once weighted by the growing base-period expenditure share, the weighted price relative for this single item reaches values as high as 1.08.

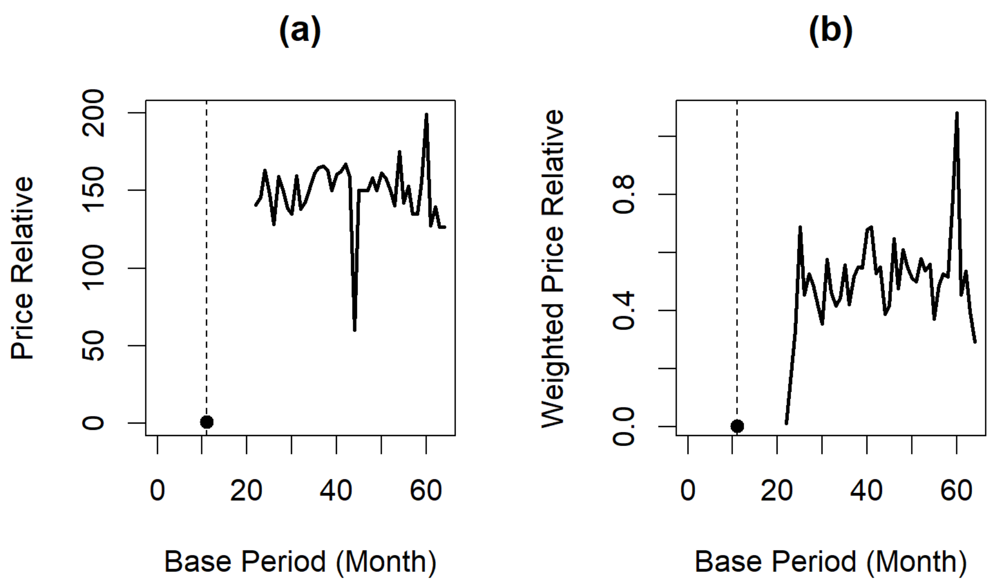

Figure 10: Price relatives (panel a) and base-period-share-weighted price relatives (panel b) for month 11 relative to each possible base period (the periods b in equation (5)), UPC 4740000000.

Figure 11 shows, as a boxplot for each base period, the distribution across every other product in the category of the corresponding base-period-share-weighted price relative for month 11. By contrast with panel (b) of Figure 10, the vast majority of these weighted price relatives are close to zero, with even the most extreme outliers among the remaining 1,380 products being well under 0.07. Hence, they are smaller by more than an order of magnitude than the values generated by UPC 4740000000 for many choices of base period. It is this combination of an extreme price relative and a non-trivial (and growing over time) expenditure weight that allows a single item to dominate the Laspeyres comparison for a substantial number of the 64 base-period comparisons with month 11 that enter the GEKS geometric mean, thereby producing the spike seen in Figure 7.

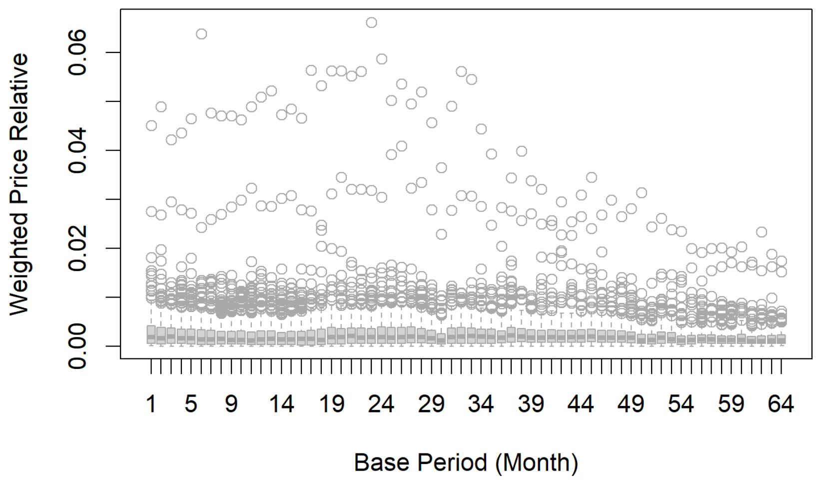

Figure 11: Boxplots of the distribution across all other products of the base-period-share-weighted price relative for month 11, by base period (UPC 4740000000 excluded; its weighted price relatives range from 0.001 to 1.081). Each box spans the first and third quartiles, with the median shown as a horizontal line. The whiskers extend to the most extreme observation lying within 1.5 times the interquartile range of the nearer quartile, and observations beyond the whiskers are plotted individually as circles.

Two features of the Törnqvist formula work together to insulate it, and hence the CCDI index, from this problem. First, the Törnqvist weight on each product is the average of its expenditure share in the two periods being compared, rather than the base-period share alone. Since UPC 4740000000 has an essentially zero share in month 11 itself (only one unit was sold), averaging its near-zero month-11 share with its (potentially larger) base-period share substantially dampens the weight it receives, regardless of which month is used as the base.[7]

Second, the Törnqvist index is a share-weighted geometric mean of the price relatives, so the *log* of the Törnqvist index is a share-weighted arithmetic mean of the *log* price relatives,

[7]Dikhanov (2025) derives a corresponding result for a general shock to an item's expenditure share.

as in equation (4). In contrast, the Laspeyres index is a share-weighted arithmetic mean of the price relatives themselves. The two functional forms respond very differently to a single extreme observation. In the Laspeyres index, an item with weight s_n and price relative r_n contributes the additive term $s_n r_n$, which grows linearly in r_n. In the Törnqvist index, the same item contributes the additive term $s_n \ln r_n$ to the *log* of the index, as in equation (4), and hence the multiplicative factor $e^{s_n \ln r_n} = r_n^{s_n}$ to the index itself. The item's price relative therefore affects the Törnqvist index only through its logarithm, and $\ln r_n$ grows far more slowly than r_n.[8]

For example, consider an item with a price relative of $r_n = 150$ and a weight of $s_n = 0.005$ in both formulas. For the Törnqvist index, whose weight is the average of the item's expenditure shares in the two periods being compared, s_n is that average; the two shares need not be equal, but need only average to 0.005. Holding the weight equal across the two formulas isolates the effect of the difference in functional form. The item then adds $s_n r_n = 0.75$ to the level of the Laspeyres index, whereas it scales the Törnqvist index by the factor $r_n^{s_n} = 150^{0.005} \approx 1.025$. As one effect is additive and the other multiplicative, comparing them requires evaluating the impact of each against a counterfactual in which the item's price relative is a non-extreme value, such as $r_n = 1$. Under this counterfactual the item would still contribute $s_n \times 1 = 0.005$ to the Laspeyres index, so the extreme observation raises the index by $s_n r_n - s_n = s_n(r_n - 1) = 0.745$. With the remaining items generating an index level in the neighbourhood of one, this is an increase of approximately 75 per cent. For the Törnqvist index, the item's factor under the counterfactual is $1^{s_n} = 1$, so the net effect of the extreme observation is to scale the index by $150^{0.005}/1 \approx 1.025$, an increase of 2.5 per cent. Unlike the Laspeyres figure, this proportional effect is exact and does not depend on the level of the index – a given additive increase represents a larger proportional change the lower is the index level. The impact on the Laspeyres index is thus around thirty times larger.[9] In the case of UPC 4740000000, the Törnqvist weight on the month-11 comparisons is in fact well below the base-period share, as the month-11 share is essentially zero. This is the first feature noted above, and it dampens the item's influence on the Törnqvist index even further.

The mechanism just derived analytically for a single illustrative item also holds up empirically, across the whole sample rather than for month 11 alone. For each of the $64 \times 63 = 4{,}032$

[8] Since Fisher is the geometric mean of Laspeyres and Paasche, and Paasche is barely affected here (Figure 8), the same mechanism produces an analogous, cubic-order, Fisher–Törnqvist gap; see Dikhanov (2025).

[9] This insulation argument applies to extremely *high* price relatives, which is the empirically relevant case here. For price relatives close to zero the ranking partially reverses, i.e. the Laspeyres contribution $s_n r_n$ is bounded below by zero, so an item can reduce the Laspeyres index by at most its weight s_n, whereas the Törnqvist factor $r_n^{s_n}$ declines without bound as $r_n \to 0$.

ordered period pairs (b, t), Figure 12 isolates UPC 4740000000's own contribution to the bilateral Fisher and Törnqvist indexes, by dividing each index as computed on the full 1,381-product category by the corresponding index computed with the item excluded, giving the item's Fisher ratio and Törnqvist ratio.[10] The figure highlights the 43 pairs comparing month 11, as target, with a base period of 22 or later – that is, the $P_{b,11}$ comparisons that directly enter the GEKS calculation for month 11 in equation (5). The reciprocal comparisons, with month 11 as the base period, are omitted. This is because both indexes satisfy the time reversal property, so that $P_F(11,t) = 1/P_F(t,11)$ and $P_T(11,t) = 1/P_T(t,11)$. Hence, the reciprocal comparisons contain no information beyond what is already shown, and are not part of the calculation that produces the documented spike. For the remaining 3,946 period pairs, in which the item either does not enter the comparison or is priced normally in both periods, both ratios are indistinguishable from 1: $|\ln P_F - \ln P_T|$ never exceeds 0.003. For the 43 highlighted pairs, the item's Fisher ratio reaches as high as 1.95, while its Törnqvist ratio stays within the much narrower range of 1.00 to 1.04. This is the same asymmetric amplification documented above for the single illustrative item, now made visible across the sample as a whole.

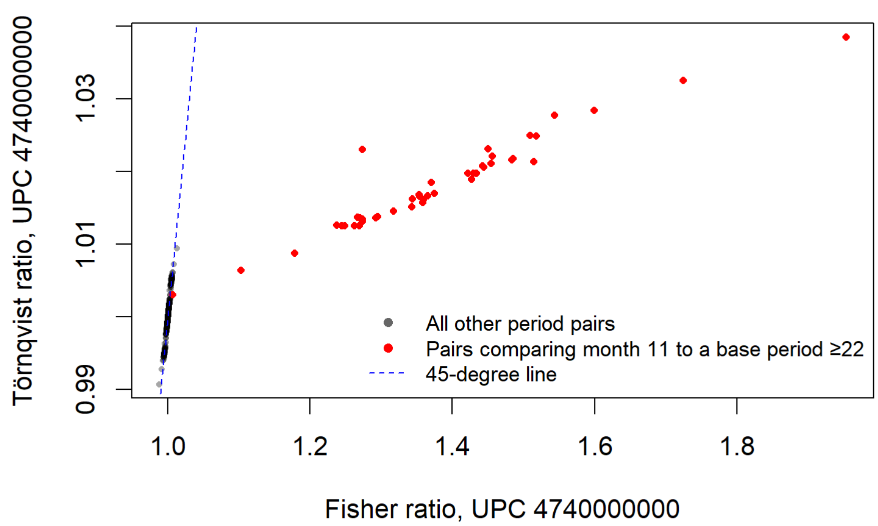

Figure 12: UPC 4740000000's Fisher ratio and Törnqvist ratio (each index as computed on the full category, divided by the corresponding index with the item excluded). Red points are the 43 pairs comparing month 11, as target, with a base period of 22 or later; the reciprocal pairs, with month 11 as base, are omitted as redundant by time reversal.

This asymmetry is not a coincidence particular to these two index formulas, or to this

[10]This nets out the item's multiplicative contribution to each comparison from the change in the price level coming from the remaining 1,380 products between the two periods concerned.

one item. Rather, it is an instance of a much more general property of weighted means. For a fixed set of price relatives and a fixed set of weights, the weighted power mean of order r is non-decreasing in r: the weighted arithmetic mean (order 1, as in Laspeyres) is at least as large as the weighted geometric mean (order 0, as in Törnqvist), which is at least as large as the weighted harmonic mean (order -1, as in Paasche), with strict inequality unless the price relatives are all equal. This is exactly the comparison made in the worked example above, where equalizing the weight s_n reduced the Laspeyres–Törnqvist comparison to one between an arithmetic and a geometric mean of the same values. Laspeyres, Paasche and Törnqvist use different weights in practice (base-period, current-period, and average shares, respectively), so the inequality does not apply to them directly. However, the underlying property carries over. That is, higher-order means respond more strongly to extreme values than lower-order means, which is why we see the Fisher index, inheriting the Laspeyres index's sensitivity to this observation, diverging from the Törnqvist index.

5 How to Contain an Explosion

Section 4 showed that using the GEKS-Törnqvist (CCDI) index in place of GEKS-Fisher is one remedy for the instability documented above, as the CCDI index is largely unaffected by the extreme observation responsible for the spike. This section considers two further remedies, for when the GEKS-Fisher is used.

Figure 7 also plots the GEKS-Fisher and CCDI indexes computed with a 25-month rolling window and mean splicing. With this window length, the GEKS-Fisher spike at month 11 is essentially eliminated, and the mean-spliced GEKS-Fisher index tracks the mean-spliced CCDI index closely throughout the sample. A sufficiently short window limits the number of base periods, within any given GEKS calculation, in which the problematic item's expenditure share has had time to grow large. This limits the extent to which it can distort the corresponding Laspeyres comparisons feeding into the geometric mean in equation (5).

Table 1 reports the month-11 value of the GEKS-Fisher index for a range of window lengths.[11] For windows of 13 and 19 months, the index at month 11 is below 1 and there is no spike at all, as the problematic item does not reappear until month 22 and hence falls outside the window entirely – as there is no other period in the window in which the item is transacted, there are no matched-model bilateral comparisons involving it. Hence, the

[11] The mean splice is used for the rolling window calculations, although, as discussed below, these values do not depend on the choice of splicing method. With a window of 64 months the whole sample forms a single window, so no splicing is involved.

small difference between the index values for these windows reflects only the different sets of products and periods entering the two windows. From 25 months onwards, the spike grows monotonically with the length of the window. At 25 months, only months 22 to 25 enter as base periods, and the item's expenditure share is still small in those months, so there is only a small effect. As the window lengthens, base periods in which the item has a substantial expenditure share are progressively included, and the spike grows accordingly, reaching 1.2186 when the full sample is used as a single window.

Table 1: GEKS-Fisher index value at month 11, by rolling window length.

Window length (months)	Index at month 11
13	0.9913
19	0.9904
25	1.0137
31	1.0564
37	1.0898
49	1.1483
64 (benchmark)	1.2186

The choice of splicing method does not matter here. With a rolling window of length W, the index values for periods 1 to W are those of the initial window, and the splicing method determines only how periods $W + 1$ onwards are appended to them. Month 11 lies within the initial window for every window length in Table 1, so its index value, along with those of all other periods within the initial window, is identical across alternative splicing methods by construction.[12] The resilience of the GEKS-Fisher index in Figure 7 to the extreme observation is thus attributable to the 25-month window length alone, and not to any property of the mean splice.

The month-11 index value thus only becomes extreme when the window used to compute it also contains the later months in which the item's expenditure share has grown. It is those months, when used as base periods, that generate the large weighted price relatives shown in Figure 10. For a period beyond the initial window, the index value is obtained from the window ending at that period, which by construction contains no later periods. Therefore, the anomaly discussed here requires the extreme observation to fall within the initial estimation window. This reflects the timing in this particular case, where the extreme price precedes the growth in the item's expenditure share. For an item whose share was

[12]The choice of splice does affect the index over later periods. For example, at month 64, the GEKS-Fisher index ranges from 1.1651 using the movement splice to 1.1827 using the WISP splice. The mean splice is used here as it is the preferred method among the available splicing methods. Evidence on the relative performance of alternative splicing methods is provided by Diewert and Fox (2022) and Fox et al. (2026).

already substantial at the time an extreme price was recorded, the relevant base periods would instead precede the extreme observation, and the same distortion could arise in any window containing both. A national statistical agency updating an index in real time would therefore not necessarily observe a distortion in the period in which an extreme observation occurs, although it may still be transmitted to subsequent periods through splicing.

This suggests that, while short window lengths can result in more chain drift bias (Fox et al., 2026), long window lengths can also be problematic. The window length must therefore be chosen with both considerations in mind.

To confirm the above analysis, all of the indexes discussed above were recomputed with UPC 4740000000 removed from the dataset. Figure 13 shows the resulting benchmark CCDI and GEKS-Fisher indexes. The GEKS-Fisher spike at month 11 has disappeared completely, and the GEKS-Fisher and CCDI benchmark indexes now track one another almost as closely as their bilateral counterparts do throughout the entire sample, exactly as the numerical-closeness-of-superlative-indexes intuition would suggest. Removing a single product out of 1,381 in the category is therefore sufficient to eliminate the entire anomaly.

This points to a third possible remedy: cleaning the data of such observations before constructing the index. Two difficulties should be noted, however. First, it requires the problematic observation to be identified in the first place, with different detection methods potentially deleting different observations. Second, an outlier is not necessarily an error, and keeping legitimate observations is important for capturing actual price movements in an index. Data cleaning is therefore best viewed as a complement to, rather than a substitute for, the use of an index method that is inherently more robust to extreme observations.

6 Conclusion

The standard result that Fisher and Törnqvist indexes approximate each other closely has been shown not to carry over to their multilateral counterparts. The multilateral GEKS-Fisher index can misbehave dramatically when applied to scanner data, producing an economically implausible spike driven entirely by a single item with an extreme, isolated price observation combined with a subsequently growing expenditure share. The GEKS-Törnqvist (CCDI) index is found to be highly resilient to this type of anomaly, involving extremely high price relatives.

This resilience derives from two features of the Törnqvist formula: (i) its weights average the expenditure shares of the two periods being compared, and (ii) price relatives enter the index only through their logarithms, which greatly limits the influence of extreme observations. The GEKS-Fisher index can also be made resilient, without abandoning the

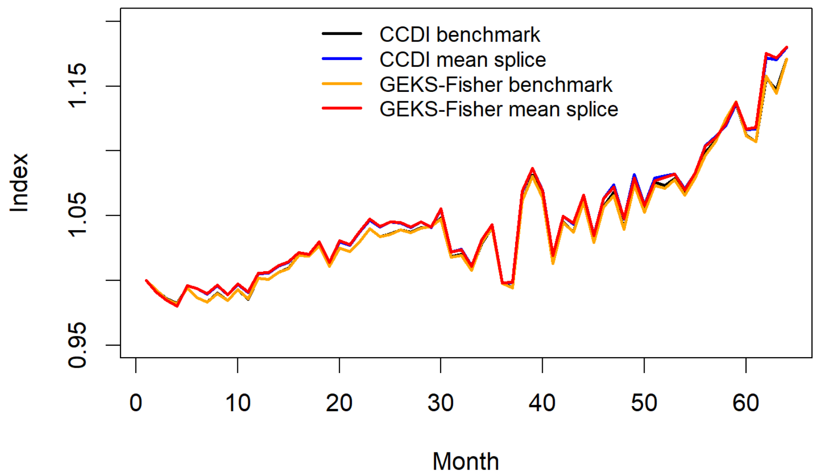

Figure 13: Benchmark and mean-spliced CCDI and GEKS-Fisher indexes after deleting UPC 4740000000.

Fisher index number formula, if it is implemented using a sufficiently short rolling window. The choice of splicing method makes no difference in this respect. Of the GEKS variants considered, the GEKS-Törnqvist (CCDI) index is preferred, due to its relative resilience to extreme observations. If the GEKS-Fisher index is used instead, considerable care needs to be taken in the choice of window length and data-quality screening.

References

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, 73–86.

de Haan, J. (2008), "Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data," paper presented at the EMG Workshop, UNSW Sydney.

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. `http://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf`

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.

Diewert, W.E. (1974), "Applications of Duality Theory," in *Frontiers of Quantitative Eco-*

nomics, Vol. II, M.D. Intriligator and D.A. Kendrick, eds., North-Holland, Amsterdam, 106–171.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics* 4, 115–145.

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

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

Diewert, W.E., N. Abe, A. Tonogi and C. Shimizu (2025), "Scanner Data and the Construction of Inter-Regional Price Indexes," *Review of Income and Wealth*, `https://doi.org/10.1111/roiw.70032`.

Dikhanov, Y. (2025), "Fisher versus Törnqvist when Relative Price Changes are Large," unpublished manuscript, World Bank.

Eltëtö, O. and P. Köves (1964), "On an Index Computation Problem in International Comparisons," *Statisztikai Szemle* 42, 507–518.

Fox, K.J., P. Levell and M. O'Connell (2026), "Inflation Measurement with High-Frequency Data," *Journal of Business and Economic Statistics* 44(1), 383–396.

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

Hill, R.J. (2006), "Superlative Index Numbers: Not All of Them Are Super," *Journal of Econometrics* 130(1), 25–43.

International Labour Organization, IMF, OECD, UNECE, Eurostat and The World Bank (2020), *Consumer Price Index Manual: Concepts and Methods*, Washington, DC: International Monetary Fund.

Inklaar, R. and W.E. Diewert (2016), "Measuring Industry Productivity and Cross-Country Convergence," *Journal of Econometrics* 191(2), 426–433.

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.

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

Melser, D. (2018), "Scanner Data Price Indexes: Addressing Some Unresolved Issues," *Journal of Business and Economic Statistics* 36(3), 516–522.

Szulc, B. (1964), "Indices for Multiregional Comparisons," *Przeglad Statystyczny* 3, 239–254.

White, G. (2021), *IndexNumR: Index Number Calculation*, R package. `https://CRAN.R-project.org/package=IndexNumR`