What do missing prices mean for the choice of index number method with alternative data?

Authors: Ben Hillman, Liam Greenhough, Alex Rose, Helen Sands

Abstract

Price missingness and product churn in scanner and web-scraped data can be problematic for the calculation of price indices using multilateral methods. Internationally, splicing methods are used in conjunction with multilateral methods to avoid revising CPIs. However, these splicing methods can be shown to downwardly bias price indices, particularly for seasonal products and those with a high churn rate. In this presentation we look at how big the problem of price missingness and product churn really is for the use of alternative data sources in CPIs, and whether price imputation could help resolve these challenges.

Introduction

The United Kingdom's Office for National Statistics is engaged in a wide range of improvements to how economic statistics are measured as laid down by the Independent Review of UK Economic Statistics by Professor Sir Charles Bean [1]. Part of the recommendations of the review included further use of scanner and web scraped data in the production of inflation indicators. However, making use of these new sources requires the development of new methods and techniques for building robust indices. ONS has published a development plan for modernising the production of consumer prices [2] which includes relevant methodological research questions such as the development of expenditure weights for web scraped prices, understanding the appropriate index methods to apply, and the effect of imputation and outliering on these data.

Traditionally price indices have been constructed using bilateral index methods using data collected and checked by price collectors [3]. However, bilateral indices struggle when there is a large degree of churn in the data that is being processed, and high churn is often a characteristic of these new, automatically collected, scanner and web scraped data sources. This can result in unrepresentative indices as entering and leaving products do not feature within the index calculations. A common solution for building index numbers from these new sources is to use multilateral indices. These indices compare two months against each other within the context of a larger "window" of months. Using multilaterals allows price statisticians to readily incorporate products that enter and leave the market, ensuring these methods better capture inflation within more dynamic markets.

Multilateral index methods were originally developed within a Purchasing Power Parities context to use with a fixed number of countries [4]. By contrast, when used in a CPI context, National Statistical Offices seek to extend their indices to each new measurement period as time goes on. To extend the time series in this way, without revising the CPI, multilateral indices are generally used in conjunction with an extension method, which can introduce some chain drift. In this paper, we explore the nature of this drift and potential solutions to the drift.

Transitory/non-transitory asymmetries

Transitory/non-transitory asymmetries explained

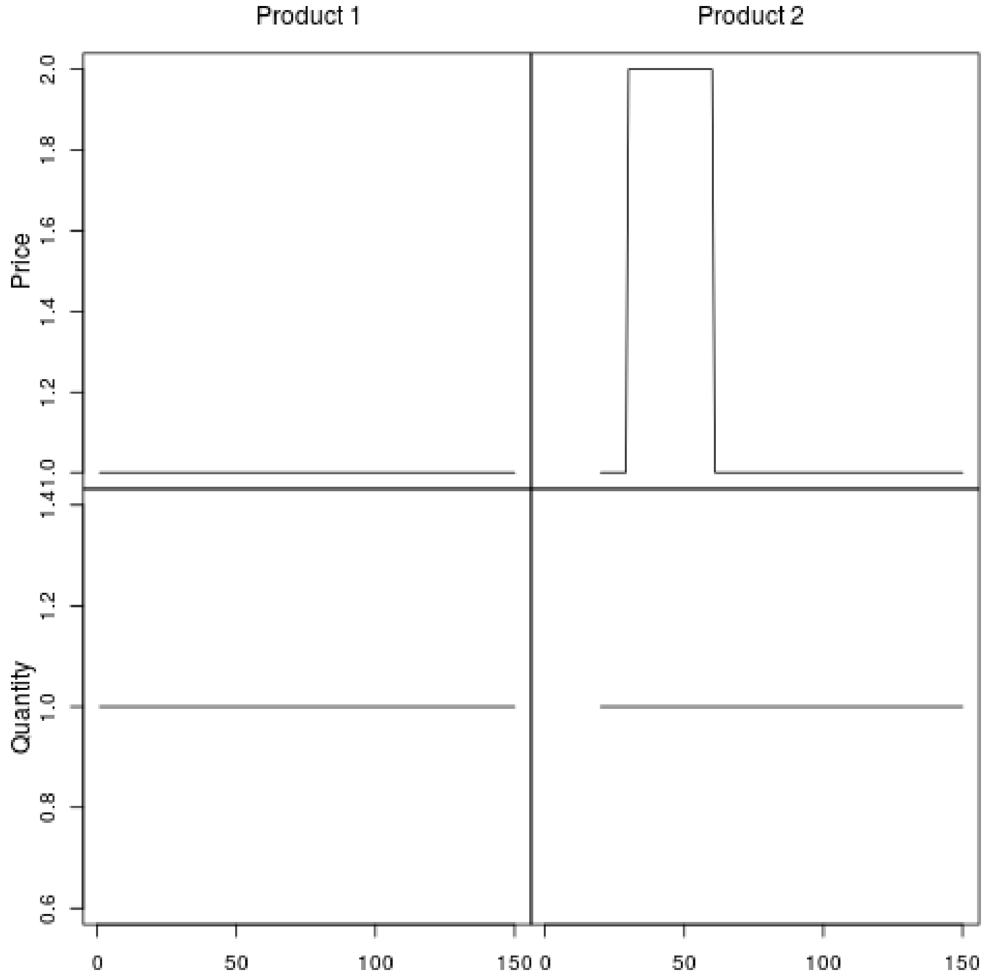
When working with new data sources, National Statistical Offices can use multilateral index methods to measure inflation from a dynamic census of evolving products. Over time, products enter the market, are around for a while, then leave the market. Since multilaterals calculate within windows, products may be fully available, partially available, or completely unavailable depending on the window being calculated. We show this, and its consequences, with an example.

Consider measuring inflation over 150 months using a 25-month window multilateral index method, in a scenario where:

- Product 1 is always available, always selling at £1 and one unit sold
- Product 2 enters the market in month 20 and…
  - Within months 20-29: sells at £1 and one unit sold
  - Within months 30-59: increases to £2 and one unit sold
  - Within months 60-150: returns to £1 and one unit sold

The price and quantity time series for these products under this scenario are shown in Figure 1.

Figure 1: Price and quantity for products 1 and 2 under our scenario



We make the following definitions within the context of multilateral calculations within a window:

- When a product is fully available for all months within a window, we describe the product as **non-transitory**
- When a product is partially available, only being available for some months within a window, we describe the product as **transitory**

For example, in Table 1, we consider two separate windows that the multilateral index method is calculated over. Month 6-30 (left) is the window where the price increase in product 2 (to £2) occurs whereas month 36-60 (right) is the window where the price return (back to £1) occurs. We note that product 1 is available in every month in both windows, so is non-transitory, whereas product 2 is transitory in month 6-30 but non-transitory in month 36-60.

In other words, there is an asymmetry: the price increase occurs whilst product 2 is transitory and it resets whilst non-transitory. We describe problems of this nature as “transitory/non-transitory asymmetries”. These asymmetries occur as products enter and leave the market.

Table 1. In the window covering months 6-30 product 2 is transitory, whereas in the window covering months 36-60, product 2 is non-transitory

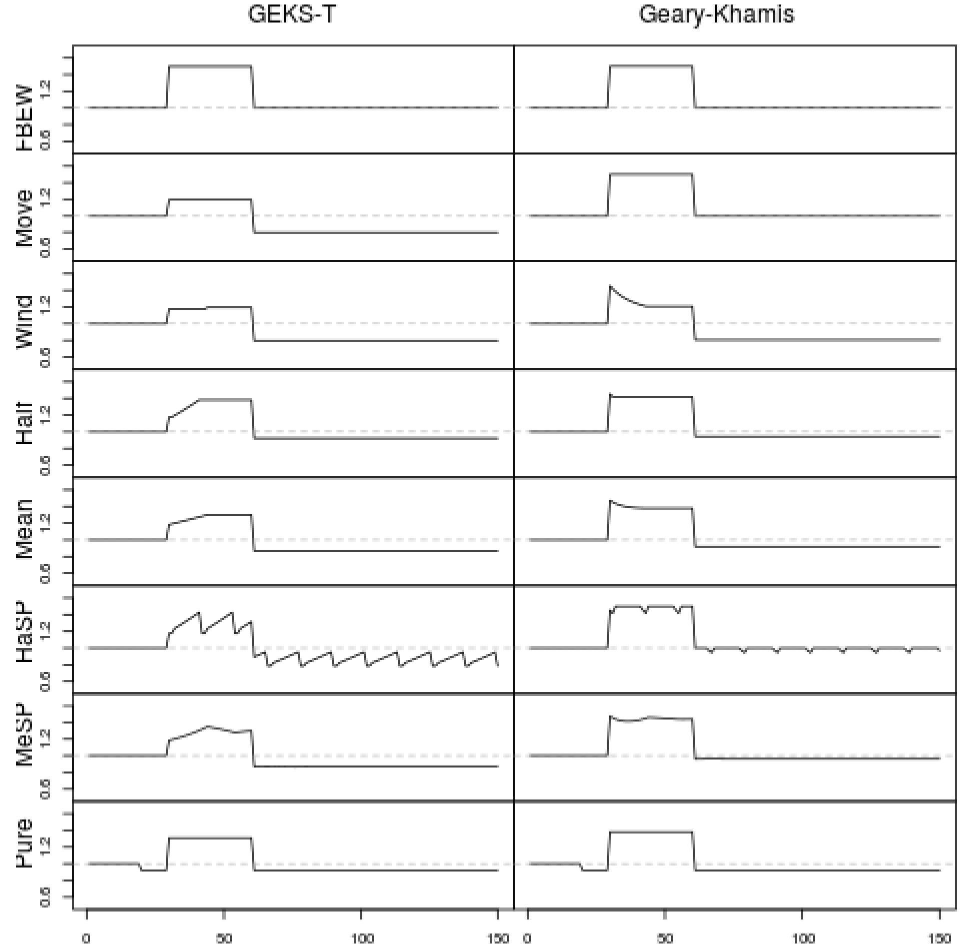
multilateral(6, 30, 6:30) product availability			multilateral(36, 60, 36:60) product availability		
Month	Product 1 is…	Product 2 is…	Month	Product 1 is…	Product 2 is…
6	Available	Unavailable	36	Available	Available
7	Available	Unavailable	37	Available	Available
8	Available	Unavailable	38	Available	Available
9	Available	Unavailable	39	Available	Available
10	Available	Unavailable	40	Available	Available
11	Available	Unavailable	41	Available	Available
12	Available	Unavailable	42	Available	Available
13	Available	Unavailable	43	Available	Available
14	Available	Unavailable	44	Available	Available
15	Available	Unavailable	45	Available	Available
16	Available	Unavailable	46	Available	Available
17	Available	Unavailable	47	Available	Available
18	Available	Unavailable	48	Available	Available
19	Available	Unavailable	49	Available	Available
20	Available	Available	50	Available	Available
21	Available	Available	51	Available	Available
22	Available	Available	52	Available	Available
23	Available	Available	53	Available	Available
24	Available	Available	54	Available	Available
25	Available	Available	55	Available	Available
26	Available	Available	56	Available	Available
27	Available	Available	57	Available	Available
28	Available	Available	58	Available	Available
29	Available	Available	59	Available	Available
30	Available	Available	60	Available	Available

Transitory/non-transitory asymmetries results

In Figure 2 we demonstrate the impact the Figure 1 asymmetry scenario has on several index method-extension method combinations (using a 25-month window). For index methods we explore the GEKS-Törnqvist and Geary-Khamis. These methods were chosen for comparisons as they are two that are widely in consideration for use by other NSOs and they are both high performing methods in the index method literature. For extension methods, we explore (in order): fixed-base expanding window; movement splice; window splice; half-splice; mean-splice; half-splice on published series; mean-splice on published series; pure[1].

[1] Note that "pure" refers to using a single extensionless 150-month window rather than a 25-month window using an extension method.

Figure 2: The following graph demonstrates the impact of the figure 1 transitory/non-transitory asymmetries on a variety of index method and extension method combinations



How well each index method-extension method combination works in the presence of transitory/non-transitory asymmetries depends on what is perceived as the "benchmark" or "goal" of the index. Product 1 stays the same price throughout the 150 months and product 2 increases in price in month 30 and resets in price at month 60, so the ideal result may be similar to how the fixed-base expanding window behaves, where the index rises and resets in line with the product 2 price changes.

If this is perceived as the goal, then the long-term index falls below 1 under many of the methods, which may not be seen as ideal. This is because multilateral index methods often give less implicit weight to products that are transitory. Product 2 is transitory when the price increase occurs and non-transitory when the price reset occurs. Therefore, less weight is given to the product in this example when its price is increasing relative to when its price is decreasing – resulting in an index that falls below 1.

The source of this implicit weighting is fairly evident within the GEKS family of methods. For example, a 25-month GEKS-Törnqvist is constructed from an aggregation of 50 Törnqvist bilateral comparisons. When a product is non-transitory, they appear in all 50 bilaterals; whereas when a product is transitory, they appear in fewer comparisons, implicitly reducing the weight of the product.

However, we must be careful not to infer too much solely using Figure 2. For example, the Geary-Khamis with a movement splice appears to capture the "price reset" well. However, this occurs when a transitory price change is followed by a non-transitory price reset. In Figure 3a, we explore the reverse, where a non-transitory price increase is followed by a transitory price reset caused by a product leaving the market. The corresponding indices shown in Figure 3b demonstrates the Geary-Khamis with a movement splice to not perform as well as in the earlier scenario. This demonstrates the need for a more thorough review of these asymmetries.

Figure 3a. In this scenario, we observe a non-transitory price increase and a transitory price reset…

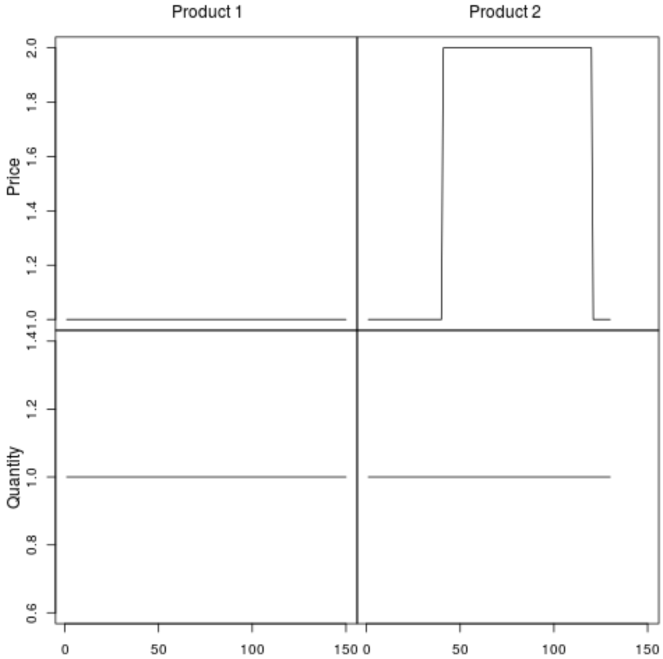
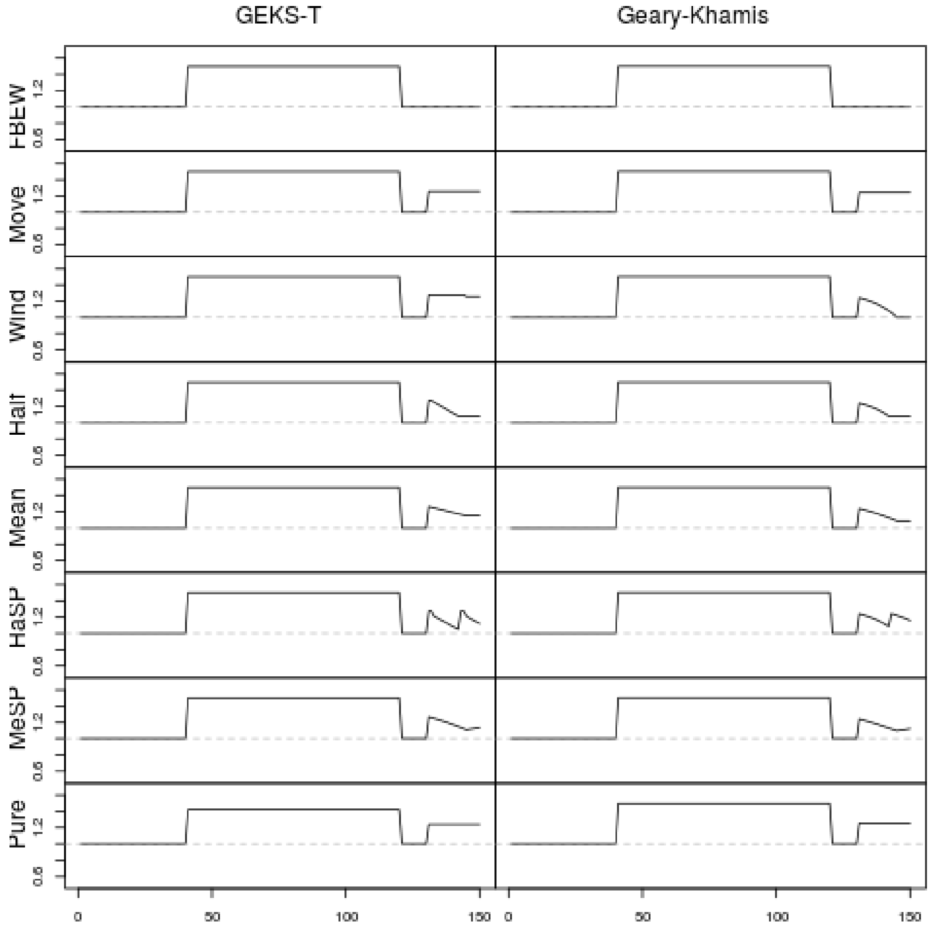


Figure 3b. … and its corresponding impact on indices



An alternative goal for the index is the “pure” series. Using this index as a benchmark for how well each of the index-extension method combinations perform changes our perception of which are the most well-performing methods at handling transitory/non-transitory asymmetries. In the next section, we consider the suitability of this benchmark.

Transitory/non-transitory asymmetries and measuring chain drift

Multilateral indices such as the GEKS-Törnqvist and the Geary-Khamis, before applying extension methods, are transitive and therefore free of chain drift. Some chain drift is introduced once extension methods are applied. Therefore, an approach for quantifying chain drift in multilateral methods (which we will explore later in the paper) involves comparing a drift-free extensionless single-window multilateral index method (which we describe as the “pure” index) against a multilateral with an extension method, which may contain some drift.

However, these transitory/non-transitory asymmetries demonstrate a curious result in the “pure” index. Note that when using a 25-month window and an extension method, the first 25 months involve calculating a 25-month “pure” index, with extension methods applied from month 26. As a result, within the first 25 months, both the pure index methods and the extension-based index methods are transitive. However, in Figure 2, we note that there is a difference between the “pure” index and the other (extension-based) indices between month 20 and 25. Since both methods are transitive within the first 25 months, we cannot describe the difference as being a result of chain drift. Therefore: what is the source of this difference?

We will consider this within the context of GEKS-Törnqvist, (although note that this difference also exists within the Geary-Khamis). Consider measuring the GEKS-Törnqvist from month 1 to month 23 within the 150-month window. The following formula is used (note that the re-arrangement holds since the Törnqvist index passes the time reversal test):

$$GEKS - Törnqvist(1, 23, 1{:}150) = \sqrt[150]{\prod_{L=1}^{150} \frac{P_{Törnqvist}^{(L,23)}}{P_{Törnqvist}^{(L,1)}}} = \sqrt[150]{\prod_{L=1}^{150} (P_{Törnqvist}^{(1,L)} \times P_{Törnqvist}^{(L,23)})}$$

Consider L in the months where the price increase in product 2 occurs, for example month 40. $P_{Törnqvist}^{(1,40)} = 1$ since it only uses product 1 whose price is unchanged and $P_{Törnqvist}^{(40,23)} < 1$ since it uses product 1 and product 2, and product 2’s price is higher in month 40 than month 23. So $P_{Törnqvist}^{(1,40)} \times P_{Törnqvist}^{(40,23)} = P_{Törnqvist}^{(40,23)} < 1$.

In other words, the pure index falls between months 20-29 by interpreting a “price fall” in product 2 from months 30-59 to months 20-29 within underlying bilaterals such as $P_{Törnqvist}^{(40,23)}$. Taking the difference between the pure index and one of the index methods with extension methods applied, in this case, does not quantify past chain drift, but is a result of incorporating future information. It may therefore seem unintuitive to consider the pure index to be a “goal”, since the pure index, in this case, is making inferences on current inflation based on price changes that have not yet occurred.

In general, it’s likely that the difference between the pure index and an extension-based index captures a mixture of chain drift and these “temporal differences”, although it is unclear how to separate the two. In the remainder of this paper, we consider using this approach to quantify this difference, although the extent to which the difference can be described as “chain drift” is still an unresolved question.

Chain drift and product churn

Data and methods

This analysis is based on the work of Kevin Fox, Peter Levell and Martin O’Connell, Multilateral index number methods for Consumer Price Statistics [5], in which they discussed and evaluated the properties of different multilateral index numbers for measuring high frequency price changes, drawing on household scanner data from Kantar.

For this analysis, we applied a similar approach to retailer scanner data held by the ONS. These data have gone through preprocessing that includes the linking of relaunched products, unit standardisation, and uses a higher level of product ID called the Shop-Keeping Unit (SKU) that groups extremely similar items together even if they have different Global Trade Item Numbers (GTINs; also known as barcodes). This preprocessing reduces the amount of artificial churn in the data.

Our analysis was also performed over each consumption segment for each region, the level at which ONS elementary aggregates will be calculated. This means that while we followed a similar approach to Fox et. al., the data used are different and the methods are applied at a different level. In Table 2 we compare some of the key differences between the Fox et al analyses and the ONS analyses.

Table 2. Data comparison of Fox et al analyses and ONS analyses

	Fox et al	ONS
Type of data	Household scanner	Retailer scanner
Relaunch linking	No	Yes
Unit standardisation	No	Yes
ID used	GTIN	SKU
Months used (number of “splices”)	84 (59)	56 (31)
# elementary aggs (con segs*regions)	52 (=52*1)	2244 (=187*12)

Note that not all 2244 elementary aggregates had enough data to calculate indices. Any time that methods are compared, only elementary aggregates where the indices could be calculated for all relevant methods were included. This means that counts will add up to less than 2244.

Fox et al suggested product churn as a potential cause of chain drift. We measure churn in the data as monthly churn, the percentage of expenditure on products in a month that were not present in the previous month. Mean churn over all periods is used to reflect the average churn for an elementary aggregate.

Our focus was on the two main methods under consideration by the ONS, the Geary-Khamis (GK) index and the GEKS-Törnqvist (GEKS-T) index [3], using the mean splice on published indices. We also considered several imputation methods, the results of two of which are included in the paper.

The roll-forwards multilateral-adjusted method fills in missing values by first calculating a multilateral index with the unmodified data first, then rolls forwards the missing prices and applies the index's change to each of the imputed prices. This is based on the inflation-adjusted roll-forwards imputation method [6].

The Time-Product Dummy method creates a log-linear model for price estimation of the form "log(price) ~ time + product-ID", with dummy variables for each time period and product, and uses this to estimate the missing prices.

Both methods were used to impute three periods forward from when a product was last sold. They were applied as 'pure' imputation methods, in that the multilateral index and the TPD model were both constructed with all available data, rather than calculated every period as would happen when applied in practice. This was done for practical reasons, as the iterative method would involve calculating an additional 67,320 indices and as many TPD models. This does mean that the imputation methods 'look forwards' and may perform better in this research than they would in practice.

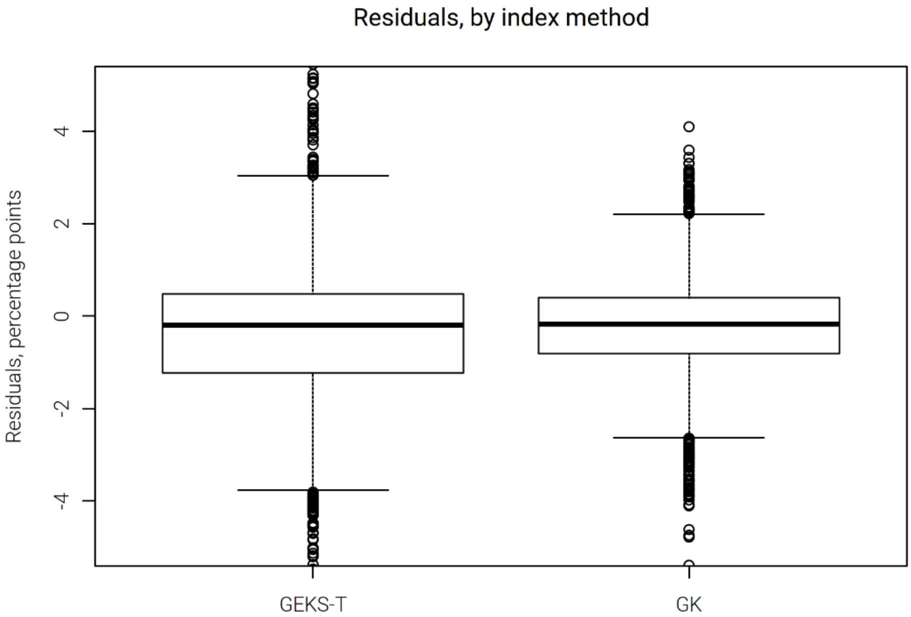
The indices calculated with the conventional methods (and any imputation) are compared to a 'benchmark' index, which is an index calculated from the same data with a pure equivalent of the same index method. This means that a GEKS-T index with a 25-month mean splice on published would be compared to a benchmark that is a pure GEKS-T index. Likewise, a GEKS-T index with 25-month mean splice on published which uses TPD imputation would be compared to a benchmark that is a pure GEKS-T index that used TPD imputation.

The difference between the index and the benchmark in the final time period (period 56 in our data) is referred to as the 'residual', calculated as index value minus benchmark value. As the benchmark index is calculated over all periods it is free from chain drift, which is why we use it as our target index. The residuals are expressed as a percentage point difference, where an index of 103 and a benchmark of 102 would give a 1 percentage point residual.

Examining the residuals

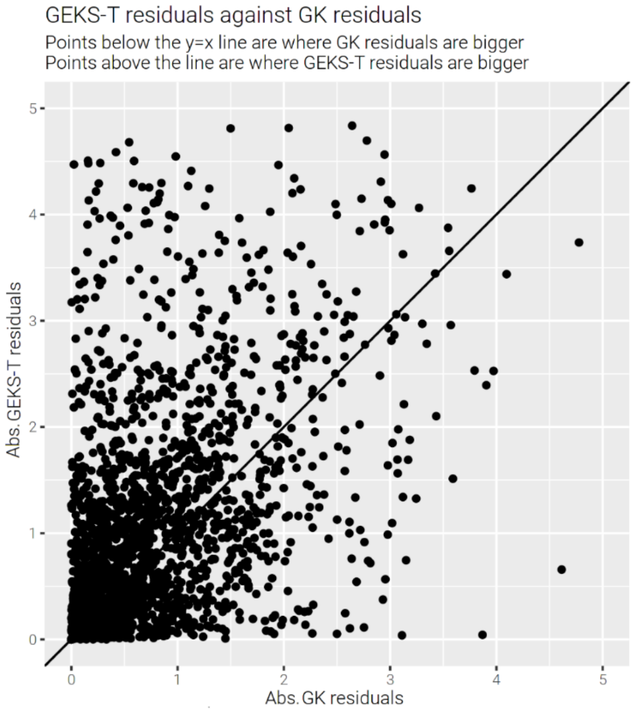
The initial results in Figure 4a, looking at the two main index methods (GEKS-T and GK [3]) without imputation, show that both the GK and the GEKS-T have the median value of their residuals very close to 0.

Figure 4a. Residuals by index method boxplot



Also visible is the marginally wider spread of residuals for the GEKS-T compared to the GK, though this hides a deeper picture of the differences between the two. Plotting the absolute GEKS-T residuals against the absolute GK residuals, as in Figure 4b, it becomes clear that an elementary aggregate that produces a large residual for the GK does not necessarily produce a large residual for the GEKS-T, and vice versa.

Figure 4b. GEKS-T residuals plotted against GK residuals scatterplot



This ties into underlying differences between the GK and the GEKS-T. As shown in the transitory and non-transitory asymmetries section, the GEKS-T may be slightly more sensitive to missing products, while Diewert & Fox's Substitution Bias in Multilateral Methods for CPI Construction using Scanner

Data [7] shows that the GK is more susceptible to the Gerschenkron effect due to how it treats substitutions.

Examining churn

To further examine these differences, we compare mean monthly churn to residuals. Results are shown for GEKS-T (Figure 5a) and GK (Figure 5b). Mean churn is log-transformed, as otherwise the points are highly clustered and hard to interpret. As a point of reference, the x value of 0 is equivalent to 1 percentage churn, where 1% of expenditure is on products new that month. We note that as the churn increases, the spread of values increases. This overall trend is the same for both the GEKS-T and GK, albeit there is a narrower spread within the GK. (Note that both graphs are to the same scale for ease of comparison).

Figure 5a. GEKS-T residuals plotted log mean monthly churn scatterplot

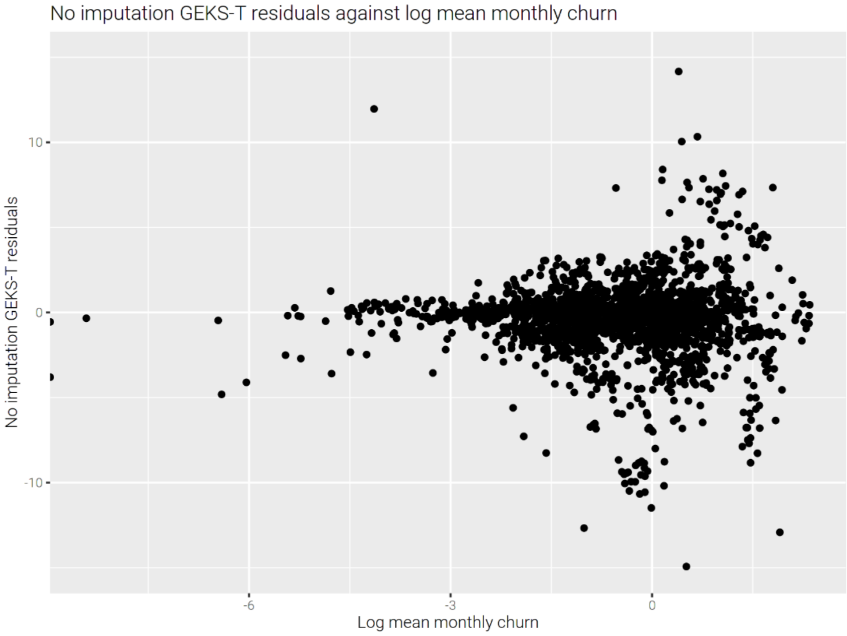
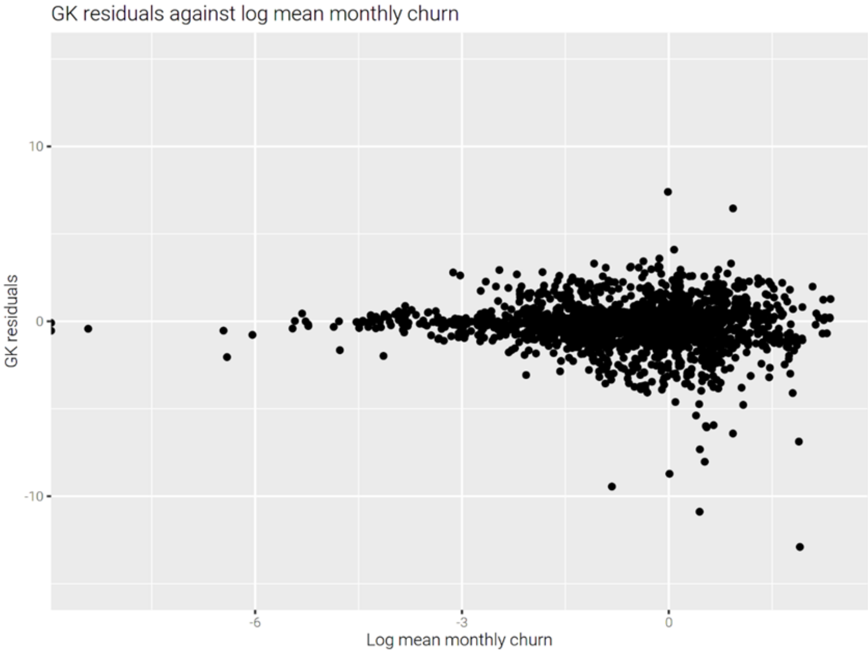


Figure 5b. GK residuals plotted log mean monthly churn scatterplot

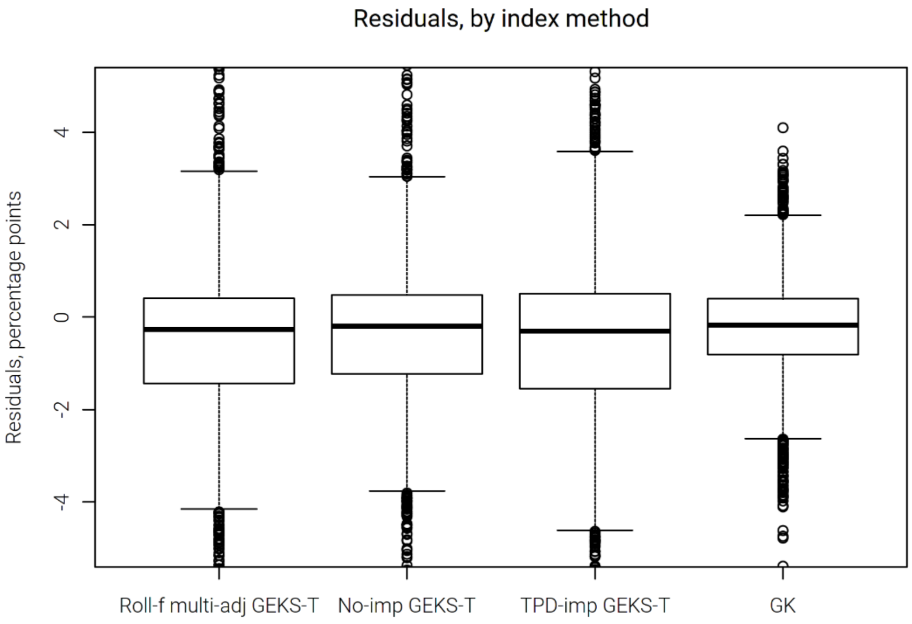


From this, increasing product churn appears to increase the size of residuals for both methods, though the GK appears less affected by it under this metric.

Can imputation solve these problems?

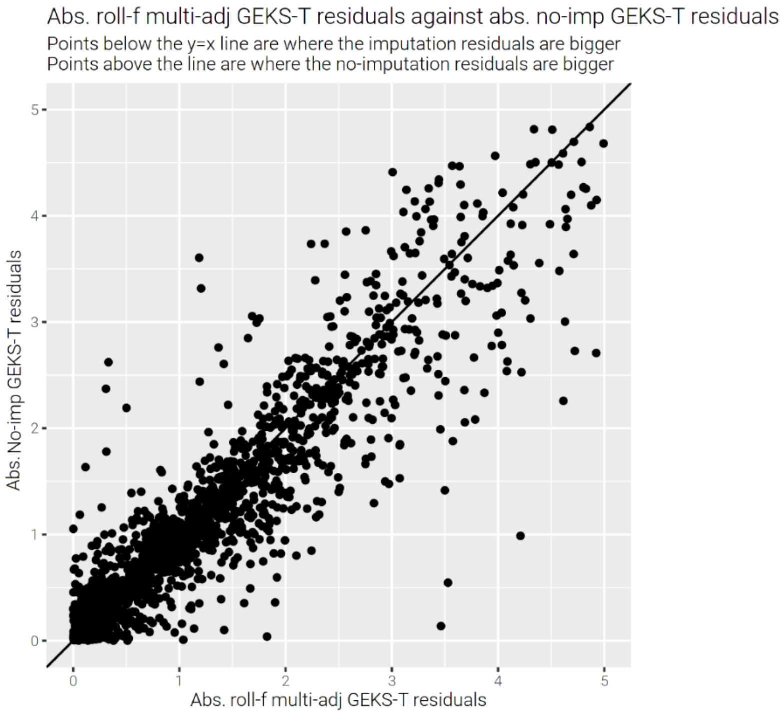
The GK index zero-weights a product with 0 quantity sold, and so is unchanged by imputation [8]. As the GEKS-T index uses bilateral Törnqvist indices, which are weighted by the average expenditure shares from the two periods, imputing the prices of products with zero quantities in only one period will change the indices, potentially reducing the effect of the missingness and the transitory non-transitory asymmetries that missingness induces.

Figure 6. Residuals by index method boxplot, now including imputation



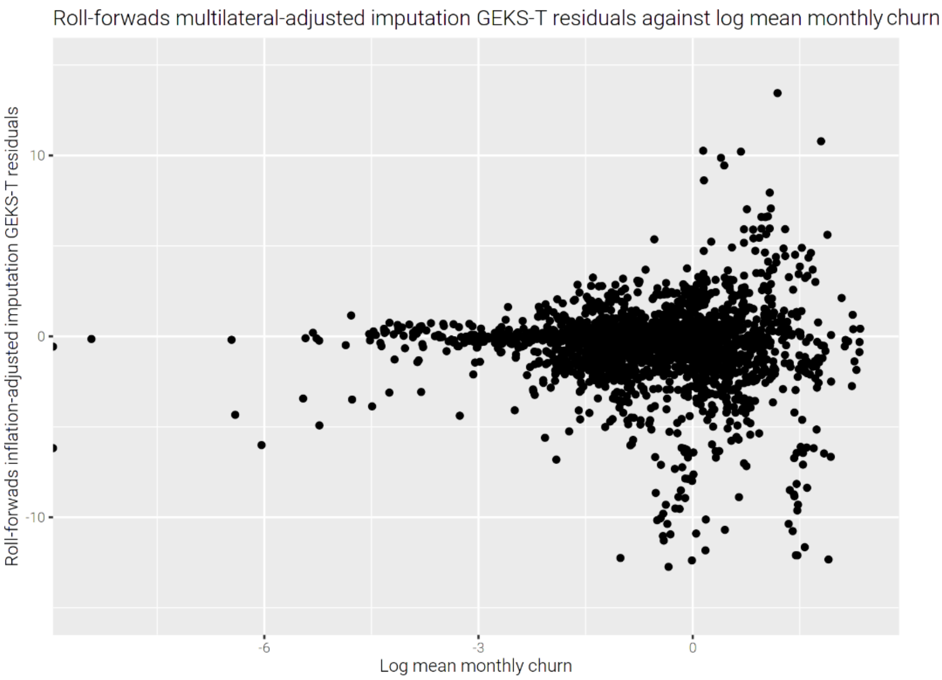
In Figure 6 we see that both imputation methods appear to produce both a wider spread of residuals, and with slightly greater median residual values, though the differences are still very small. When examining the roll-forwards multilateral-adjusted imputation index's residuals against the no-imputation index's residuals (Figure 7a), as expected there is a clear relationship between the two. If the imputation had reduced the magnitude of the residuals, then we would expect to see more points falling above the line, but with 922 points above the line compared to the 1204 below the roll-forwards multilateral-adjusted imputation does not appear to improve performance by this metric.

Figure 7a. GEKS-T residuals plotted against GEKS-T with absolute roll-forwards multilateral-adjusted imputation residuals scatterplot



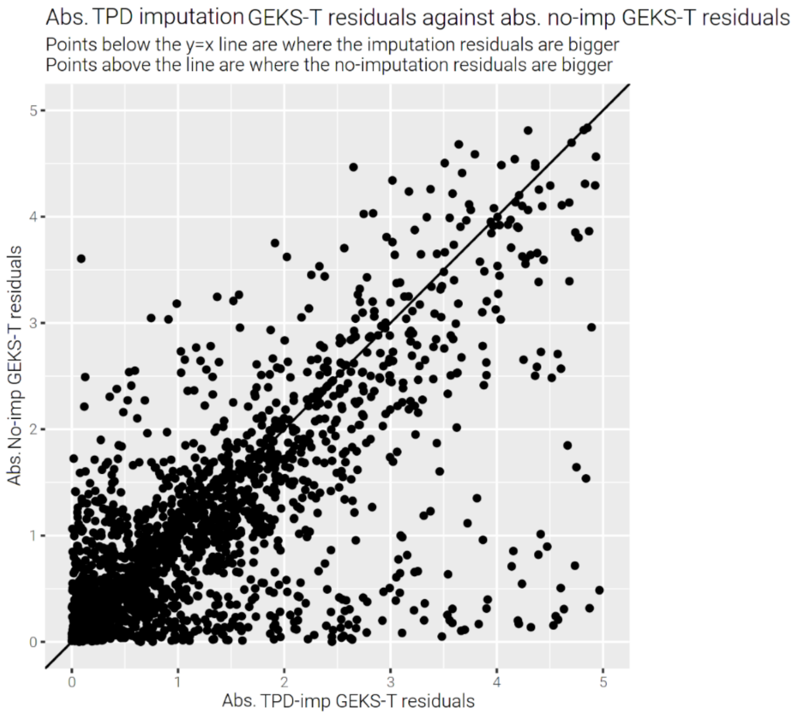
When examining the effect of this imputation on the residuals against churn graph (Figure 7b), the increasing churn still corresponds to an increase in magnitude in the residuals, which does suggest that the imputation hasn't resolved the issue and may have in fact made it worse.

Figure 7b. GEKS-T with absolute roll-forwards multilateral-adjusted residuals plotted against log mean monthly churn scatterplot



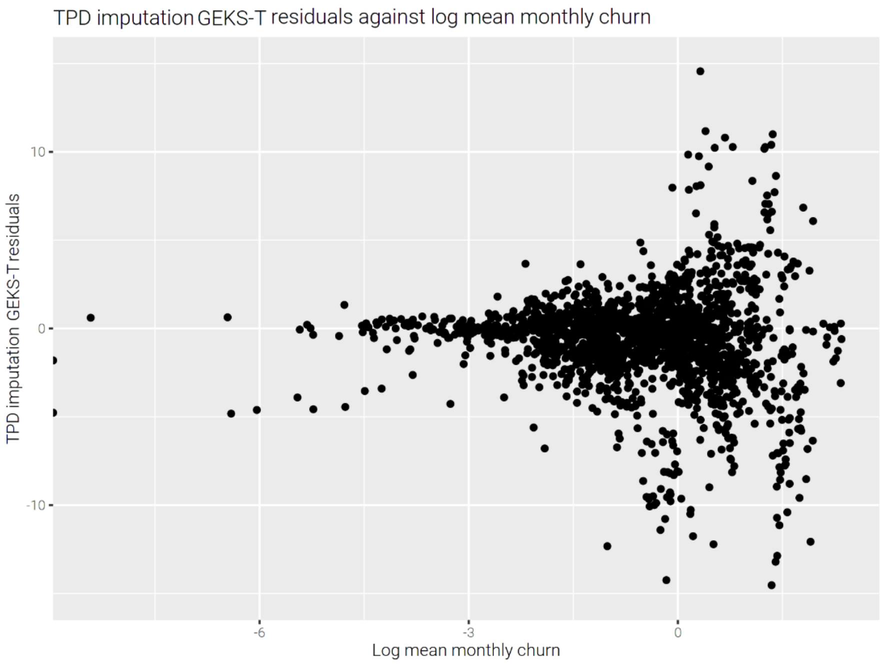
As shown in Figure 8a, the TPD-imputation GEKS-T also has a similar ratio of points above the line to points below; at 933 above and 1193 below, but the graph reveals additional detail. There are a noticeable number of points, running just above the x axis, where the TPD-imputation GEKS-T shows substantially larger residuals with very low comparable residuals for the no-imputation GEKS-T, which is a more clear-cut example of imputation making the problem worse.

Figure 8a. GEKS-T residuals plotted against GEKS-T with absolute TPD-imputation residuals scatterplot



Comparing the TPD-imputation residuals to the churn (Figure 8b), the general pattern of increasing magnitude of residuals as churn increases is still apparent, meaning that by the above measure TPD-imputation does not correct for missing prices in practice.

Figure 8b. GEKS-T with absolute TPD-imputation residuals plotted against log mean monthly churn scatterplot



Summary and future work

Our work indicates that product churn is at least correlated with the magnitude of residuals. This suggests that higher churn will have a negative impact on all index methods. Further, our work indicates that the imputation methods examined do not appear to resolve this issue.

We only imputed over three periods, a very conservative limit, and it might be that imputing over more periods (up to 24 for a product only present in a single period of the 25-period window) would reduce the transitory/non-transitory asymmetry effect, but also runs the risk of unrepresentative prices being included in the index. As the GEKS index considers all periods within a window equally, imputing over all periods within a window, and doing this imputation separately for each window, may be in keeping with the idea of the method even if it is an unreasonable level of imputation under traditional methods.

The measure of churn used, monthly churn, is also not the only metric; annual churn (percentage of expenditure on products that month that aren't present in the previous 12 months) and weak seasonality (measure of products which are always available but have regular fluctuations in prices or quantities that are synchronized with the season or the time of the year [9]).

The main caveat for this paper relates to the benchmark method used. The principle behind using a pure index is that it experiences no chain drift, but due to the nature of multilateral methods a pure index will 'look ahead', which can cause unexpected and counter-intuitive indices. Pure indices also include more time periods than windowed and spliced indices, and there is space for further work to isolate the chain drift effect from the change in characteristicity of longer multilateral indices.

References

[1] Independent Review of UK Economic Statistics, Bean, C. (2016)

[2] Consumer prices development plan, Payne, C., ONS (2021)

[3] New index number methods in consumer price statistics, Sands, H., ONS (2020)

[4] Price and Quantity Index Numbers – Models for Measuring Aggregate Change and Difference, Balk, B. (2008)

[5] Multilateral index number methods for Consumer Price Statistics, Fox, K., Levell, P., and O'Connell, M. (2022), ESCoE Discussion Paper 2022-08.

[6] Scanner Data, Elementary Price Indexes and the Chain Drift Problem, Diewert (2018a)

[7] Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data, Diewert, W. & Fox, K., Journal of Economic Literature Classification Numbers (2017)

[8] Scanner Data in the CPI: The Imputation CCDI Index Revisited, Haan and Dalmaas (2019)

[9] Comsumer Price Index Manual: Theory and Practice, chapter 22 – Treatment of Seasonal products (2004)