ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Index methods: framework for using multilaterals

Status: Final
Expected publication: Alongside the minutes for October 2022

Purpose

1. In this paper, we present a range of lower-level index methods choices that we need to make in preparation for introducing alternative data sources for second-hand cars and rail fares in 2023. We consider:
    a. recapping how alternative and traditional data sources are aggregated together
    b. how multilaterals/alternative data sources will work within an annual round framework
    c. how we pre-emptively detect and filter out index calculations in months where the GEKS-Törnqvist cannot be calculated due to missing data
    d. how we impute to fill these filtered-out months

Actions

2. Members of the Panel are invited to:
    a. advise whether the methods outlined in this paper are suitable
    b. express any objections to use of "the partial GEKS" and "month-on-month aggregate imputation" for imputation purposes

Background

3. This work is a continuation of a series of papers taken to APCP seeking to finalise decisions on ONS index methodology when applied to alternative data sources. In our previous papers taken to APCP:
    a. April 2022: we presented our decision to use GEKS-Törnqvist with a 25 month window and a mean splice on published series as the preferred multilateral design choices with alternative data sources, following a review by Fox, Levell and O'Connell and our own internal analyses.
    b. July 2022: we presented guidance to the Panels on how and why these new methods work in practice, with the intention of publishing these materials in late-2022.
4. We will incorporate the multilateral combination outlined above, within the framework outlined in the remainder of this paper, when introducing alternative data sources for rail fares and second-hand cars in 2023.

Aggregation (a reminder of the key points)

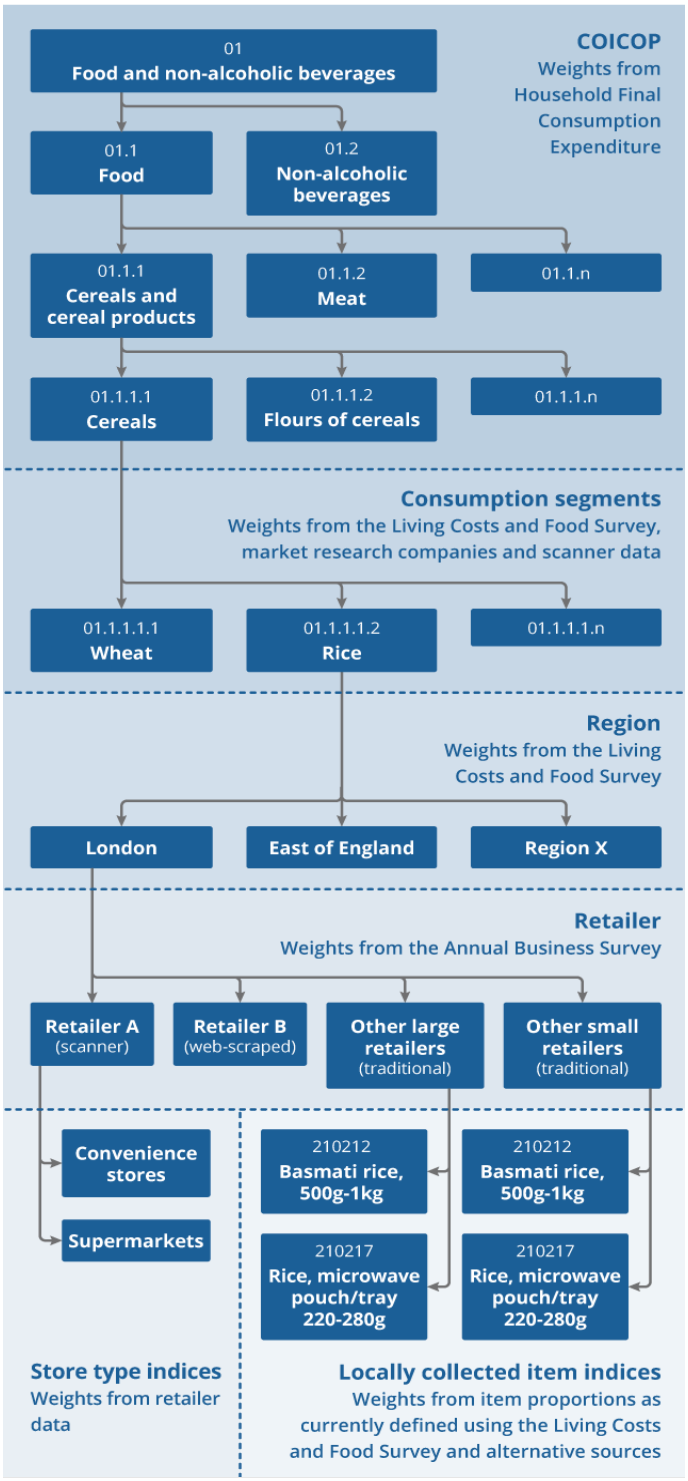
5. In Figure 1, we show the aggregation structure we have presented in our previous aggregation publication. We recap the key points:
    a. consumption segments are introduced as a level below COICOP.
    b. a region level is introduced to calculate regional inflation more readily in future.
    c. alternative data sources are stratified by retailer, while traditional data are stratified by large (>2% market share) and small (<2% market share) retailers.

d. alternative data sources may contain additional strata below retailer level, depending on the category. Elementary aggregates for ADS cover all appropriate products fitting the pre-specified strata within the consumption segment.

e. traditional data sources continue to use items, which are representative of the consumption segment (for example, in figure 1, basmati and microwave rice represent "rice"). Product samples are then chosen for items.

f. we continue to use Lowe aggregation above the elementary level.

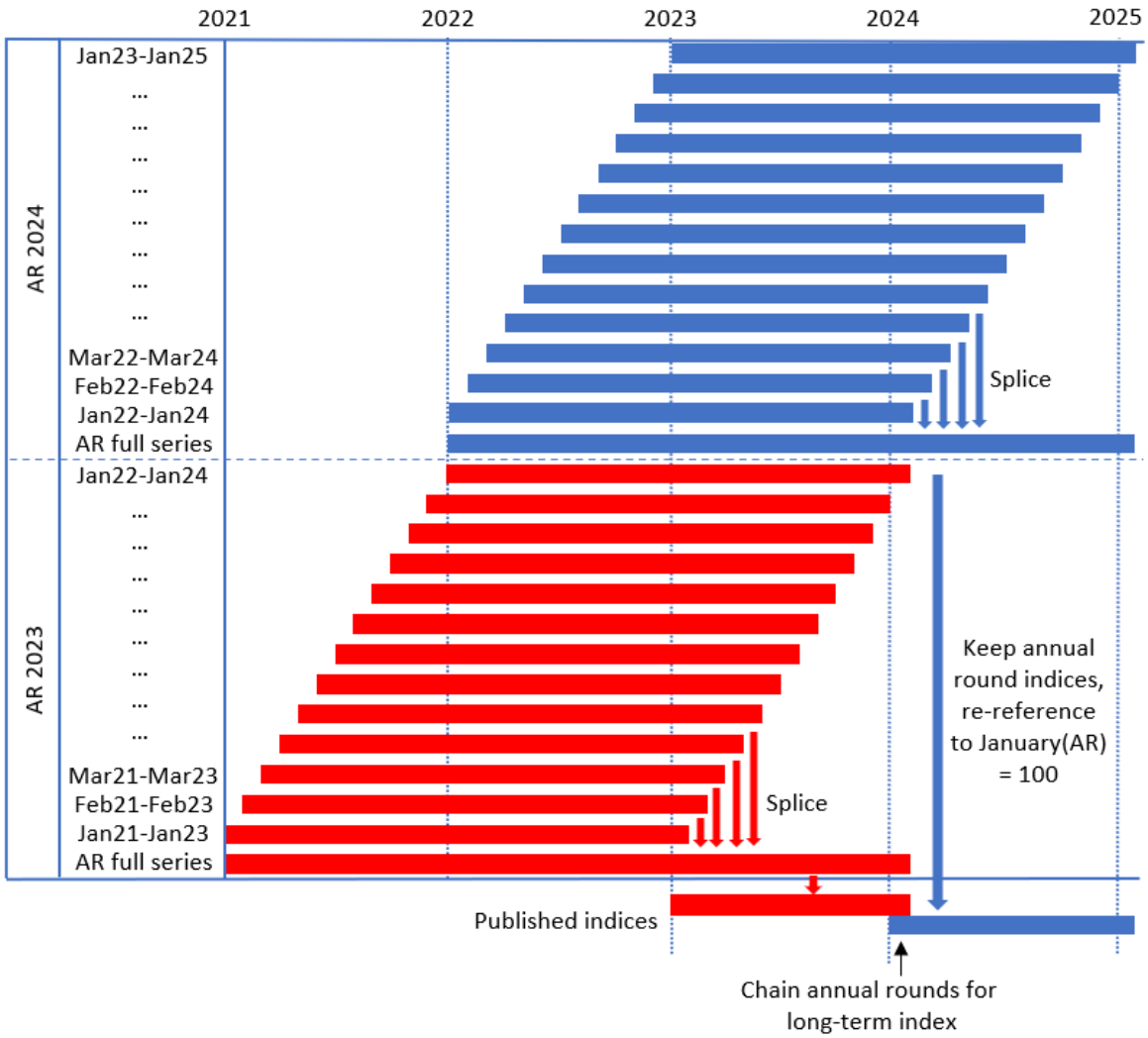
Figure 1: The planned aggregation structure to integrate alternative/traditional data sources



Alternative data sources within an annual round framework

6. We will continue working within an annual round framework, providing the opportunity to make structural changes to our consumer price statistics each year, including updating weights, updating the basket items or consumption segments to reflect changes in consumption, and allowing us to introduce alternative data sources for new categories or retailers.
7. Aside from the new aggregation structure, the annual round process for traditional data sources remains broadly unchanged.
8. For alternative data sources, where we will use multilaterals, we need an approach for producing indices with annual January "breaks" to incorporate these structural changes. Figure 2 shows how we do this.

Figure 2: Using multilateral index methods within an annual round (AR)



9. For the 2023 annual round (covering January 2023 to January 2024, the bottom-half of figure 2), we:
    a. initialise our "annual round full series" by calculating a 25-month GEKS-Törnqvist window covering January 2021 to January 2023 for each elementary aggregate (this allows us to introduce new consumption segments/ retailers etc).
    b. each month, roll the window on by one month, calculate the GEKS-Törnqvist within the new 25-month window and splice[1] a new month onto the annual round full series.
    c. Subset the annual round published series to only the months covering the annual round, re-referencing to January(y) = 100 so that they are on the same basis as all other data sources ready for aggregation.
10. The above process is repeated within the 2024 annual round (covering January 2024 to January 2025). To obtain the long-term series for an elementary aggregate from 2023 to 2025, we single chain-link these two annual rounds together using January as a shared month.

Incalculable GEKS-Törnqvist filters

11. The GEKS-Törnqvist is calculated as an aggregation of many bilateral Törnqvist indices. When the component bilateral Törnqvists are unable to calculate, this can lead to the GEKS-Törnqvist failing to produce an output for an elementary aggregate. We have therefore created checks to pre-emptively determine whether the GEKS-Törnqvist can or cannot be calculated:
    a. The "all-months product check" determines whether there is at least one unique product with price>0 and quantity>0 in every month of the window. If true, then bilateral Törnqvists can be calculated between any two given months within the window, which ensures that the GEKS-Törnqvist can be calculated.
    b. The "Törnqvist check" independently checks every Törnqvist used to calculate the GEKS-Törnqvist can be calculated.
12. The "all-months product check" is sufficient but not necessary for determining whether the GEKS-Törnqvist is calculable. That is, if this check fails, the GEKS-Törnqvist index may still be calculable. However, the "all-months product check" is less computationally expensive than the "Törnqvist check" and most indices pass this check, so this is done for expediency.
13. The "Törnqvist check" can be used if the "all-months product check" fails, and independently checks whether each individual Törnqvist can be calculated. Suppose we wanted to calculate the GEKS-Törnqvist from month 1 to month 23 with a 25-month window. Table 1a shows how we calculate the GEKS-Törnqvist and Table 1b shows how we check each of these individual Törnqvists.

[1] The method we use to splice works methodologically in the same way as what is internationally known as the "mean splice on published series". The key difference is that we do not splice directly onto what is published, but rather on the "annual round full series" that keeps the aggregate entirely consistent within the annual round.

Table 1a: GEKS-Törnqvist calculation table

L	Törnqvist(1,L)	Törnqvist(L,23)	product
1	Törnqvist(1,1)	Törnqvist(1,23)	Törnqvist(1,1)*Törnqvist(1,23)
2	Törnqvist(1,2)	Törnqvist(2,23)	Törnqvist(1,2)*Törnqvist(2,23)
3	Törnqvist(1,3)	Törnqvist(3,23)	Törnqvist(1,3)*Törnqvist(3,23)
…	…	…	…
24	Törnqvist(1,24)	Törnqvist(24,23)	Törnqvist(1,24)*Törnqvist(24,23)
25	Törnqvist(1,25)	Törnqvist(25,23)	Törnqvist(1,25)*Törnqvist(25,23)
		GEKS-Törnqvist(1,23) =	**geometric average(product column)**

Table 1b: In the "Törnqvist check" we check which component Törnqvist indices are calculable

L	Törnqvist(1, L) calculable?	Törnqvist(L, 23) calculable?	both_true?
1	TRUE	TRUE	TRUE
2	TRUE	TRUE	TRUE
3	TRUE	TRUE	TRUE
…	…	…	…
24	TRUE	FALSE	FALSE
25	TRUE	TRUE	TRUE

14. To determine whether the GEKS-Törnqvist is calculable, we count the "both_true" column in Table 1b. This tells us how many of the rows within Table 1a can be used within the geometric average. There are three cases:
    a. If count(both_true) = 25, then the GEKS-Törnqvist can be calculated normally, using all pairs of Törnqvists.
    b. If count(both_true) = 0, then the GEKS-Törnqvist is incalculable, and imputation is unavoidable.
    c. If 25 > count(both_true) > 0, then only some of the pairs of Törnqvists are incalculable. We then have two options:
        i. Calculate the GEKS-Törnqvist as a geometric average of the available rows (using a "partial GEKS").
        ii. Imputing the aggregate in-line with our imputation strategy (not using a "partial GEKS").
15. We are particularly looking for feedback from the Panel on the appropriateness of using a "partial GEKS". On the one hand, Eurostat (PDF, page 24) advise that the "partial GEKS" does not guarantee transitivity. By reducing the number of components in the averaging, some additional volatility may be introduced. On the other hand, full transitivity is already lost through splicing and high-level aggregate imputation and using a "partial GEKS" ensures inflation is based on genuine observable data rather than imputations.
16. In particular, we are looking to use the partial GEKS to account for rare scenarios where there is an empty product set for an elementary aggregate within a month (either due to a data supply issue or a product shortage). The partial GEKS avoids the need for "consistently imputing" beyond that month. For example, if there were a shortage of kiwis within March 2023, then not only would this mean the March 2023 index was incalculable, but also March

2023 could not be used as a link month in the GEKS-Törnqvist calculations for April onwards, leading to needing to impute every month after March, regardless of whether products were back in stock from April. Partial GEKS allows March to be eliminated as a link month, enabling the continued use of a non-imputed index.

Month-on-month aggregate imputation for elementary aggregate indices calculated using alternative data sources

17. In the previous section we outlined an approach for filtering out incalculable indices. If we choose to use the "partial GEKS" then the need for imputation may be partially avoided, but even still there may be cases where none of the underlying Törnqvists are able to calculate and the need for imputation exists.

18. Consider Table 2. We are working within the 2023 annual round with a 25-month window that is initialised between January 2021 and January 2023. We have already published February 2023 and now wish to extend our time series into March 2023. To do this we splice our new window (March 2021 to March 2023) onto the "annual round full series" using the overlapping 24 months covering March 2021 to February 2023 (although not done here, we then re-reference to January 2023 = 100 and only publish the 2023 indices as part of the 2023 annual round, as discussed in the annual round section).

Table 2: Window indices calculated with filtered out "gaps" that need to be imputed

Month	Annual round full series (pre-splice)	New window to be spliced
Jan-21	100.00	
Feb-21	100.39	
Mar-21	100.43	100.00
Apr-21	100.59	99.94
May-21	100.78	100.24
Jun-21	100.86	FILTERED OUT
Jul-21	101.37	100.25
…	…	…
Feb-23	107.64	105.37
Mar-23		FILTERED OUT

19. In Table 2 there are two months within the March 2021-March 2023 window that have been filtered out and now require imputation. To achieve this, we perform month-on-month aggregate imputation. This is done in two ways, dependent on which month fails (given as the examples):

    a. June 2021 (applies to month 2-24 within the window): calculated by applying the June month-on-month growth rate from the annual round full series to the window May 2021 index: $June\ index = 1.0024 \times (1.0086/1.0078)$

    b. March 2023 (applies to month 25 within the window): calculated by applying the March month-on-month rate from the consumption segment to the February 2023 published index directly (avoiding splicing)

Why month-on-month aggregate imputation

20. In the previous section we outlined how we use month-on-month aggregate imputation. There are two alternatives we could also consider:
    a. Year-on-year aggregate imputation works similarly to month-on-month but uses an annual instead of monthly adjustment
    b. Product-level imputation imputes prices at the product-level. The method we have explored is inflation-adjusted roll forwards imputation, where missing products are imputed using a month-on-month inflation adjustment.
21. Year-on-year aggregate imputation is a less pragmatic solution when considering elementary aggregate indices using alternative data sources. In Table 2, we cannot use year-on-year to impute June 2021 using information within the annual round since June 2020 is undefined. Instead, we would need to retrieve information from the previous annual round, which is not always possible, particularly if the retailer/consumption segment is new to the annual round.
22. Aggregate imputation (month-on-month and year-on-year) isolates the effect of imputation solely to the months that need to be imputed. Product-level imputation, on the other hand, affects indices across the entire window. Imputing product prices in June 2021 affects not only June 2021, but also every other month within the window (since other months use June 2021 as a link month).

Why month-on-month aggregate imputation: grocery case study

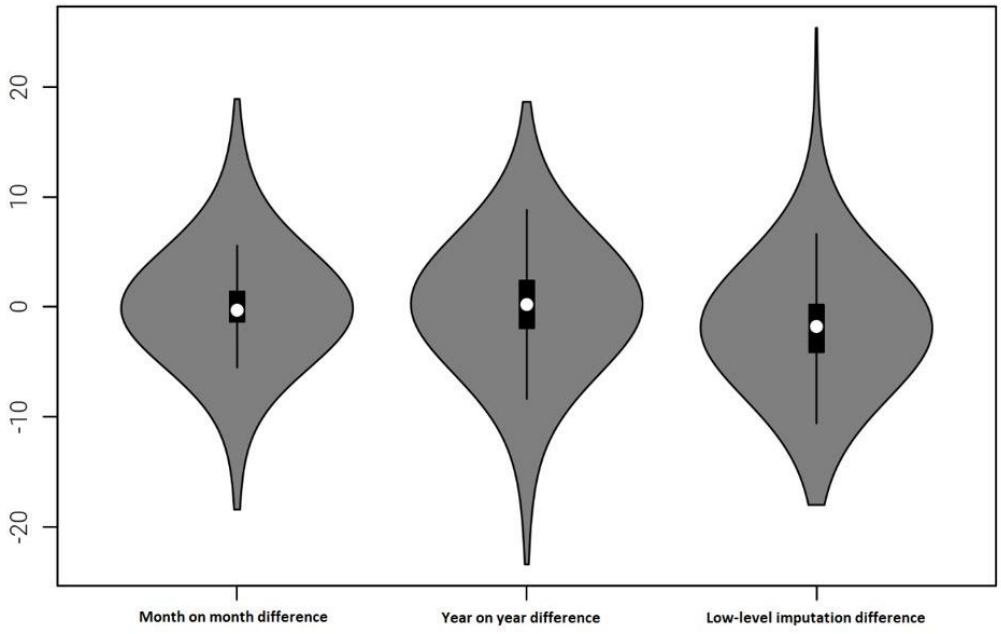
23. There is a philosophical argument on what the goal of imputation should be. One goal is for the imputed aggregate to have a "neutral" effect on higher-level aggregates, where month-on-month aggregate imputation and year-on-year aggregate imputation preserves the monthly and annual rates (when calculated without the missing index) respectively.
24. A possible second goal is for the imputed index to be as close to the index value that would have existed had the information been available to calculate the index. Note that this target is generally unknown in practice and therefore may not be seen as the primary goal. However, in this section we have explored how close the imputation methods get to a "true value" when we artificially introduce gaps into the data.
25. As a case study, we have explored the impact of applying these different imputation techniques on a single grocery retailer. There are 131 consumption segments, 12 regions and one shop type used, for a total of (131*12*1=) 1,572 elementary aggregates used in the analysis.
26. We note that imputation is expected to be rare. Calculating all 1,572 elementary aggregates over three years (mimicking an annual round), we found only 13 elementary aggregates needed to be imputed in one or more months:
    a. 12 of these elementary aggregates were the regional variants of offal
    b. The other elementary aggregate was cigars in Northern Ireland
27. We have analysed the impact of the choice of imputation. We calculated the GEKS-Törnqvist with a 25-month window and a mean splice on published series with January 2022 as our reference month and March 2022 the final measurement month, with data covering January 2020 to March 2022 used for the analysis, consistent with the annual round approach described earlier. These index series were used as our "ground truth" benchmarks. We dropped the offal elementary aggregates since a ground truth could not be measured for

these elementary aggregates, giving us a total of 1,560 elementary aggregates used in the analysis.

28. We removed March 2022 and imputed using the three approaches. Each elementary aggregate therefore has a benchmark index series, along with an index series for each of the three imputation methods. We take differences (imputed index minus benchmark index) within March 2022 to determine the quality of the imputation method.

29. In Figure 3, we created violin plots and associated boxplots comparing the benchmark to each of the three methods, over all 1,560 elementary aggregates. In our results, month-on-month aggregate imputation appears to be the most consistent with the benchmark, although all three approaches appear to perform reasonably well. Inflation-adjusted roll forwards imputation appears to be producing a small downwards bias, but this may be specific to the month chosen rather than a general observation.

Figure 3: Difference in final-month indices: imputed series (x-axis) minus benchmark



30. In Figure 3, we imputed all 1,560 elementary aggregates. However, as mentioned previously, imputation is expected to be rare. To explore the impact of imputation on higher-level aggregates we use a probabilistic sampling approach to introduce missingness. We use two scenarios, where elementary aggregates are given a 2% and 10% chance of being dropped and imputed in March 2022. We aggregated the resulting series and take differences in the higher-level aggregates between the imputed and benchmark series. Table 3 shows the results.

31. As can be seen, with a base index of 100, both the month-on-month and year-on-year aggregate imputation techniques produce very similar results to the ground truth series. Month-on-month imputation appears to be preferred overall, with results closer to our calculated all-groceries index and better results in the alcohol/tobacco division, but year-on-year imputation appears to give slightly preferable results in the food and non-alcoholic drinks division.

Table 3: In scenarios where elementary aggregates had 2% and 10% missingness/imputation rates, imputation has a very small effect on headline indices (base index = 100)

COICOP	(MoM imp. - benchmark)	(YoY imp. - benchmark)
Scenario 1: 2% Missingness		
All-groceries	-0.00230	-0.01845
(1) Food and non-alcoholic drinks	0.01168	-0.00527
(2) Alcoholic drinks and tobacco	-0.03690	-0.05107
Scenario 2: 10% missingness		
All-groceries	-0.00402	-0.12908
(1) Food and non-alcoholic drinks	0.03966	0.03312
(2) Alcoholic drinks and tobacco	-0.11212	-0.53051

32. Note that a limitation of this analysis is that every elementary aggregate is given an equal chance of being dropped. However, in practice it seems more likely that elementary aggregates with lower product counts/expenditure shares (e.g. offal) are more likely to be dropped than elementary aggregates with larger product counts/expenditure shares (e.g. chocolate). This may mean imputation has an even more limited effect than suggested in Table 3.

33. In summary, month-on-month aggregate imputation appears to be preferred for its simplicity, practicality, the fact it preserves the month-on-month rate, and there is a mild preference for it based on our analysis if the goal is to estimate a "true" value. We recommend it as our preferred method for introducing alternative data sources. However, further work on imputation is discussed in the next section.

34. Note in related work, we have also published an Ottawa paper in part exploring whether product-level imputation can solve for chain drift in multilaterals.

Second-hand cars: hybrid and electric cars

35. A decision in second-hand cars, that has been made between our earlier research publications and our current impact analysis, is to not yet go into production with consumption segment indices for hybrid and electric cars. The reason for this is that had we introduced hybrid and electric cars, we would have had to perform perhaps unacceptable levels of imputation.

36. Our original intention was to publish hybrid and electric cars as a combined aggregate. There were seven age strata and 23 car make strata that were in scope for a total of 161 elementary aggregates that would be aggregated into a combined hybrid/electric second-hand car consumption segment.

37. We attempted to produce indices using a GEKS-Törnqvist over a 25-month window covering January 2020 to January 2022. When calculating the 25th month within this window, January 2022, we noted that of the 161 elementary aggregates:

    a. 54 elementary aggregates could not be calculated.

    b. 49 elementary aggregates had at least 10 missing Törnqvist indices (out of the 47 non-trivial bilaterals).

    c. 18 elementary aggregates had between one and nine missing Törnqvist indices.

    d. 40 elementary aggregates can be calculated normally.

38. Note that these figures only cover calculating indices over a single 25-month window, but an annual round would involve calculating and splicing together 13 distinct 25-month windows. So, within an annual round we would expect even more imputation than has been outlined above.

39. Additionally, this assumes that we produce a single hybrid and electric car consumption segment but, following discussion with market experts at Auto Trader, we have concluded that these markets are not homogeneous enough to be considered within the same consumption segment. However, if we create two separate consumption segments for hybrids and electrics, then we have even further reduced samples within elementary aggregates and we would expect an unacceptable level of imputation required.

40. Therefore, we have recommended not producing consumption segment indices for hybrid and electric second-hand cars at present until a point in future where the sample sizes are robust enough to produce meaningful indices with limited missingness.

Further work on imputation

41. Our empirical analysis has primarily focussed on groceries. We expect imputation to be extremely rare for rail fares, since we do not expect product churn (for example, provided one person travels from Bradford to Leeds on a single ticket every month, then the Yorkshire single ticket will be calculable). However, due to the granularity and sometimes infrequent sales within the rarer strata of second-hand cars, some imputation may be expected (as seen for hybrid and electric second-hand cars).

42. Longer-term, we may re-consider the use of product-level imputation in the context of other goods categories. Within a 25-month window, the GEKS implicitly weights products that are ever-present higher than products that are only available in some of the months. In categories such as clothing, where the lifecycle of a product is extremely short, this may mean the index exaggerating the economic importance of the relatively few products that have extremely long lifecycles. Product-level imputation may therefore be preferred for these higher-churn categories.

Liam Greenhough and Ben Hillman
Prices Division/Methodology
October 2022