ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Shortlisting appropriate index methods for use on web scraped and scanner data

Purpose

1. In the absence of consensus amongst statistical organisations about the best choice of index method for use on web scraped and scanner data, an assessment of methods against predetermined criteria has been developed to reduce the number of possible methods for use by ONS.

Actions

2. Members of the Panel are invited to:
    a) comment on the criteria used to assess the appropriateness of methods in this paper
    b) comment on the methods chosen as appropriate for ONS to use
    c) advise on areas for future work

Alexander Rose
Methodology, Office of National Statistics
May 2019

List of Annexes

Annex A	**Literature review**
Annex B	**A summary of price index number methods**

Introduction

Scope

3. The Office for National Statistics (ONS) is investigating the use of new data sources to ensure it provides improved statistics and utilises more efficient ways of working. Globally, National Statistic Institutes (NSIs) in recent years have recognised the impact of big data in the production of statistics. Big data makes it possible to paint a more complete picture because it provides more information on a certain topic, which leads to higher confidence in the data and in turn the statistics it produces. For consumer price statistics, web scraped, and scanner data are the big data sources which are being most widely used in NSIs. This paper aims to complement existing manuals and achieve the following outcomes:

    - Summarise available literature/methods for compiling price indices using web scraped or scanner data;
    - Bring attention to existing practice across NSIs in the use of alternative data sources for price indices;
    - Indicate which methods may be appropriate for web scraped and scanner data at the ONS.

4. This paper is not intended to provide recommendations on methods for price index compilation, but instead to form ideas about the appropriateness of methods when applying these to alternative data sources. The next stage of the work will produce a framework for deciding the appropriate index methods to use for different web scraped or scanner data sets; investigating different features that may be found within these new data sets, such as high churn, high price variance or clearance prices. The final goal of this workstream is to provide recommendations on what methods are suitable for incorporation in this production stage. These methods will be used to calculate experimental price indices over the period 2020 to 2023 to allow time for suitable impact analysis. It is hoped that the framework will be generalisable enough that when new categories are introduced beyond January 2023, the same guidelines can be applied to decide which methods can be used for these categories. The comprehensive workstream is can be found in APCP-S(19)04 Alternative data sources roadmap.

5. The index methods described in this paper will be primarily of use at the elementary aggregate (EA) level. EA price indices are aggregated into item and then COICOP level indices with methods that are chosen based upon regulations; EU member states are regulated to use Laspeyres type indices in the production of their HICPs (the UK's CPI).

Background

6. Opportunities to access and analyse big data sets have become more frequently available, increasing the potential of statistical organisations to deliver new insights and improve existing output statistics. Because of the complexity and the increase of usable variables found in big data sets, NSIs face a challenge to implement more efficient and innovative approaches. As such, the ONS, like many NSIs, is undertaking research into use of big data sources for the compilation of consumer price statistics.

7. Currently the two most prominent alternative data sources for consumer prices compilation purposes are scanner and web scraped data. Scanner data is large in volume and usually

contains detailed information about sales transactions, including; product descriptions, quantities, expenditure, dates, and turnover of products sold. Web scraping is used to extract large amounts of data from websites, notably however, this data rarely provides quantity information. Web scraped prices can be collected as frequently as required using automated programs that find the related information and store it as a time series. Potentially, scanner and web scraped data can be used to improve the accuracy of current price statistics and reduce the associated costs of data collection.

8. The upper levels of most national consumer prices are aggregated using the Lowe formula, a "Laspeyres type" index method. The lowest point at which the Lowe formula is applied is known as an elementary aggregate (EA). At the EA level, NSIs have a variety of formulas which they can apply; the methods section of this paper discusses these formulas. Indices at the EA level are not currently published at ONS. However, it is worth noting that some indices at EA are equal to the indices at item level which are published.

9. The most appropriate index formula to apply is dependent upon the data source used. An index formula that is appropriate for scanner data may not be appropriate for field collected prices. Consequently, internationally the incorporation of scanner or web scraped data typically involves the creation of new retailer specific EAs. This allows for the application of appropriate weights to the different data sources.

10. The use of scanner data for weighting means that these weighting patterns can be created specifically for each item for each retailer, and for each region; provided the data provides this geographical dimension. However, scanner data expenditure may not always align with web scraped or field collected data, so adjustments may be needed. With web scraped data an external source of data or methodology for applying a proxy is required to calculate appropriately weighted EAs, due to the lack of expenditure information provided.

Motivations for using alternative data sources

11. The traditional approach to construct consumer price statistics uses probability proportional to size sampling (PPS), where retailers are sampled on an estimated market share of a group of products. Field officers attend a selected retailer and will record prices and any applied discounts or special offers. Whilst this is a recognized approach for consumer price collection, conducting personal visits is potentially costly when compared with the cost of implementing a system that would utilise scanner or web scraped data. Contrary to this is the possibility, that some item's prices will unlikely ever be collected by alternative data sources and for those that can be covered with scanner or web scraped data, may need further scrutinising to ensure the quality, this would impact any overall financial benefit.

12. Along with the possible decrease in costs of physically collecting data, the alternative data sources also mean a large-scale increase in the sample size and a likely reduction in provider burden; whilst prices will no longer be collected manually, the provider would be required to send their data to ONS.

13. Scanner data will usually contain the full inventory of products sold by a retailer, therefore improving the representativeness of the sample and reducing the probability of sampling error. Therefore, the accuracy of the index could be enhanced. Another advantage of scanner data for creating price statistics is the availability of timely quantity information.

14. Web scraped data has similar advantages to those of scanner data in terms of increased product sample size and the timeliness of the information collected. The full inventory of products from retailers' websites can be scraped at a very high frequency; daily or even hourly, at a low cost to a statistical agency. Currently ONS do not web scrape internally for consumer prices and contract an external company to provide, in general, weekly data.

Drawbacks to alternative data sources

15. Both scanner and web scraped data may be vulnerable to both over coverage and under coverage. For example, scanner data may contain business expenditure resulting in over coverage, and web scraped data may miss small retailers with no website, resulting in under coverage.

16. Bentley and Krsinich (2017) highlights that we should be aware of the data source itself; "Big data (transaction, online and administrative data) is 'found data' in the sense that measuring CPI inflation is a secondary use of the data – the data was not created with this use in mind". For example, product descriptions are generally not as detailed as NSI may want them to be because the retailers do not need them to be.

17. Another practical issue for web scraped data is that web site structures are updated frequently, this means that scrapers must be adapted to cope with any changes, which can be time consuming and can lead to missing data while the scrapers are being fixed. Scrapers also need to deal with the dynamic pricing found on websites, and ensure the scraped prices are representative of the first price presented to the consumer, not a price that has been affected by cookies, or the location of the consumers IP address. Websites can also be prone to changes in terms and conditions, so they need to be checked regularly, to ensure legal use any data the scraper collects.

18. One drawback of web scraped data by comparison with scanner data is the absence of quantity information, preventing the explicit weighting of products relative to their economic importance.

19. Whilst these limitations are recognized internationally, there are new methods and technology trialled by NSIs that mean that these drawbacks can be mitigated, and NSIs have not stopped in their pursuit of research and effective implementation of these new data sources.

Data Composition

20. Table 1 provides a summary of the expected metadata ONS receives from scanner and web scraped data sets, although this will change depending on individual retailers and websites.

Table 1: Representative file structures of scanner and web scraped data

Scanner data			Web scraped data		
Variable	**Format**	**Example**	**Variable**	**Format**	**Example**
Date of Sale	Date or Numeric	20181207, 07/12/2018	*Date*	Date or Numeric	20181207, 07/12/2018
Retailer ID	Text or numeric	Shopname, 01001	*Retailer ID*	Text or numeric	Shopname, 01001
Region of Sale	Text or numeric	Manchester			

Product identifier	Numeric or text	ABC_123	*Product identifier*	Numeric or text	ABC_123
Retailer classifications	Text or numeric	Grocery – Alcohol – Wine – Red Wine	*Retailer classifications*	Text or numeric	Grocery – Alcohol – Wine – Red Wine
Description	Text	'Classy' Pinot Noir Red Wine 75cl	*Description*	Text	'Classy' Pinot Noir Red Wine 75cl
Quantity	Numeric	100 units			
Turnover	Numeric or currency	£600.00			
			Price	Numeric or currency	£12.95

Considerations for calculating price indices

21. To ensure the use of alternative data sources are methodologically suitable for the construction of price indices there are several areas to consider. As well as deciding which index formula to use and, for multilateral methods, which is the most appropriate extension method, NSI's must also consider other areas that will impact each price index, such as:

- How a product is defined (individual GTINs or groups of GTINs?);
- How to implement index aggregation structures using different data sources (including classification);
- The choice of product weights;
- The length of the time window (multilateral methods).

22. The importance of product definition can be highlighted as follows; traditionally NSI will incorporate the COICOP hierarchy in the aggregation structure with EA the absolute lowest level at which price indices are calculated, lying below publication level. The item level is the lowest level that ONS publish and is aggregated to the COICOP level. For example, the item "Men's shirts", will typically contain GTINs with short and long-sleeved shirts, of differing colours and could contain many or one shirt per pack.

23. If we were to class all shirts as the same quality in a single unit value for each period, this produces a unit value index output for the price index at the item level. However, this would not be an accurate reflection of price change for that item. For example, if consumers bought more expensive silk shirts in a month, this would produce an increasing price index. This would be inaccurate as the shift in behaviour is one of quality instead of price. However, we must also consider a drawback of the alternate approach, not grouping GTINs, and defining each GTIN as a unique product. For certain COICOP groups, this risks a downward bias, for example when a product relaunches the "hidden" price increase will be missed if we just follow the individual GTIN. Therefore, a balance between the two extremes may be a suitable approach in some instances, where the product groups are still defined enough that unit value bias will be less of an issue, but also that product relaunches and subsequent price increases can be captured in the index.

24. It is necessary to consider issues relating to homogeneity and the implementation of index aggregation structures alongside a decision on a suitable price index that uses alternative data sources. These issues are particularly relevant for scanner data as it is usually in a disaggregated form. Summarising this issue; “at the first stage aggregation, when we are constructing vectors of prices and quantities for two periods in order to insert these vectors into a bilateral index number formula, we are forced to aggregate the individual transactions which occur within a period into some sort of period average prices and total quantities” (Ivancic et al. 2011).

25. With scanner data, a unit value for a product (either defined as an individual GTIN or a broader homogenous group) can be attained by dividing the sum of its total expenditure (value) by the sum of its quantity. When scanner data is available from individual outlets, unit values can be calculated for each separate location. This is desirable when the service offered at these locations differs, as this would account for any change in quality. However, studies have shown that prices set across all outlets of a store are nearly uniform and this suggests that it is appropriate to calculate a unit value price for a product at the retailer level rather than at the outlet level. Aggregating to the retailer level results in a reduction in the number of records making it easier for the retailer to supply the data as well as reducing the processing burden for the NSI. On the other hand, receiving the data in an already aggregated format can create data quality issues that would have been exposed if the data was disaggregated. NSI should also be aware that if outlets are free to set their own prices, then scanner data should be supplied at the outlet, rather than retailer level.

26. In general, national consumer price statistics are published monthly, but online prices may be scraped far more frequently, with some instances hourly or daily. Ideally a unit value should include information from all days of the month, for web scraped data the arithmetic or geometric average of collected prices should be calculated to arrive at a monthly price, currently ONS has chosen to use the geometric average. For scanner data a monthly unit value can be calculated by dividing total monthly sales by total monthly quantities sold. There is also a question as to how to produce a monthly unit value if product groups are introduced; should average prices of all individual GTINs first be calculated then take an average of averages; or should an average be taken of all prices together, which implicitly gives more weight to products that appear more often over the month.

27. It is also important to know how alternative data will map to the classification structure used for consumer prices. Products need to be mapped to the publication level at a minimum. Retailer classifications are usually more detailed than the published level, so mapping is made easier and many NSI choose to use the retailer’s specific categories within their own aggregation structure. However, these specific categories may not fit within the official classification and retailers may revisit and change their own classifications, imposing further burdens. Van der Grient and de Haan (2011), have shown another significant obstacle when using a retailers classification system; “chains sometimes group items related to special occasions like children’s birthdays. Products that serve a particular aim, such as meat, charcoal and sauces bought when organizing a barbeque, may also be grouped together”. These groups of products cannot be assigned to one COICOP category. Methods are available to overcome these issues, for example some NSI use machine learning techniques for classification.

28. Another important decision is in the choice of product level expenditure weights. Comparisons between weighted and unweighted methods, highlight that differences can be

significant at EA and COICOP levels, exposing the benefits of choosing a method that uses weights at product level. However, the chosen method should also be free of chain drift. This was a dilemma that has complicated method choice in previous years, leading to the increased use of multilateral methods which can satisfy both these properties. Chessa (2016b)

29. NSIs must also recognise the potential impact of window length when using multilateral methods. The choice of length has generally defaulted to 13 months or 5 quarters, ensuring the time frame accounts for any seasonality in a product's prices. The comparisons in Chessa et al. (2017 and 2017b) and Chessa (2019) shows that the differences between windows of 13 months and 4 years are small, and even negligible in many cases. Stats Netherlands uses a 13-month window for every type of product, with the GK method. ABS prefer to use a window of 9 quarters for their grocery data and Statistics New Zealand uses an 8-year window for rental prices. This is an area of research still under investigation by the statistical community, and an ideal window length has not been concluded.

Assessment of methods against criteria

30. An online literature review was conducted to inform the criteria for considering what makes an index number method appropriate when using alternative data sources. The findings of this literature review are found in **Annex A** – Literature review.

31. **Annex B** - a summary of price index number methods presents eight bilateral methods; Lowe (Laspeyres type); Paasche; Fischer; Törnqvist; Unit Value; Jevons; Dutot; and Carli, which can be extended by chain linking or using the fixed base extension method, and six multilateral methods; GEKS-J; GEKS-T; GEKS-F; Geary Khamis; Time product dummy; and Time dummy, with six different methods for extending the index without revising it.

32. In the absence of consensus amongst NSIs about the best choice of methods found in **Annex A**, an assessment of these methods against predetermined criteria has been developed to reduce the number of possible methods, selecting the most appropriate to cover the various parts of the consumer basket that can be covered by alternative data sources in future. This shortened list of methods will then be used in a follow-on paper determining a framework which highlights appropriate methods for given data set characteristics and specific areas of the consumer basket, as part of the ONS drive to make greater use of alternative data sources.

33. A natural starting point for the assessment criteria is ONS guidelines for measuring statistical quality, which outlines best practice for measuring and reporting on the statistical quality of outputs. This promotes five dimensions of statistical quality:

- Relevance – The degree to which statistics meet current and potential needs of the users.
- Accuracy and reliability – The closeness between results and the true value.
- Timeliness and punctuality - Refers to how quickly and frequently the statistic is published.
- Accessibility and clarity - Accessibility is the ease with which users can access the data. It also relates to the format(s) in which the data are available and the availability of supporting information. Clarity is the extent to which easily

comprehensible metadata are available, where these metadata are necessary to give a full understanding of the statistical data.

- Coherence and comparability - Coherence is the degree to which the statistical processes, by which two or more outputs are generated, use the same concepts and harmonised methods. Comparability is the degree to which data can be compared over time, region or another domain.

The next section uses these in conjunction with the criteria used by ABS set out in Table 3 to assess the methods.

34. **Resources**
The resources criteria would fall into the timeliness dimension of the ONS quality criteria and aims to answer the question “does this method enable more effective use of human and information resources?”.

35. This paper considers methods with the aim of making the most of alternative data in the ONS suite of consumer prices. These methods also offer the chance for automating routine manual processes and improving the use of human resources, making it more viable to produce more frequent outputs. The opportunity to automate processes to reduce costs associated with producing consumer prices is attractive to all NSIs.

36. In terms of computing requirements and processing power needed, bilateral methods will be less intensive than multilateral methods as they require less periods of data to compute the statistical output. Methods that use expenditure data (weighted methods) and/or product characteristics (hedonic methods), will also place higher burden on information resources due to the increase of variables used in the computation of the index. New data of any kind usually requires new procedures and utilising alternative data could require an increase in resources at the newly created modules of the CPI process. For example; the process of classifying products to the associated COICOP hierarchy is required every time new data is secured. These classification models will also need to be maintained over time to ensure new products entering the market are classified correctly. This mapping is currently quite a labour-intensive process as each respondent can have unique classification structures. To improve the mapping process, NSIs including the ONS are investigating the automation of the mapping process which will reduce the burden on resources and facilitate increased use of alternative data.

37. However, an increase of information resources requirements for any method could be countered by a decrease in human resources when using web scraped or scanner data. Increasing the number of price observations from alternative data sources and using a multilateral method to produce item level indices would reduce human resource requirements at various stages of the production process. Sampling, collection and scrutiny of prices all become less resource intensive. Procedures that require less manual involvement should be completed closer to the reference period and make the production of higher-frequency outputs more feasible.

38. **Theoretical properties**
This criterion belongs to the accuracy and reliability dimension. The theoretical properties of some index number methods have previously been investigated by ONS in the Mayhew (2018) paper, “ONS methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data”.

39. In index number theory there are several ways to assess which methodology is suitable, three of those approaches used by Mayhew were:

    1. the axiomatic/test approach (see International Labour Organisation (ILO) consumer price index manual chapter 16) – the index is tested against some desirable properties;

    2. the economic approach (see ILO consumer price index manual chapter 18) – the index is ranked against whether or not it approximates or is exact for a Cost of Living index, however the ONS' target is not defined;

    3. the statistical (Stochastic) approach (see ILO consumer price index manual chapter 16) – each price change is an observation of population value of inflation and the index is the point estimate of inflation.

40. ILO (2004) evaluates bilateral price indexes both from axiomatic/test and economic approaches. The methods that are recognised as the most appropriate are the Fisher and Törnqvist indexes. Under the axiomatic approach only the Fixed Based Jevons index passed all the axioms. However, it is unsuitable to use a fixed base method for a monthly production index using alternative data sources because it relies on products existing in every month and these alternative data sources will have greater product coverage and higher levels of churn. The GEKS-J and the Chained Bilateral Jevons index also performed well and pass all axioms other than price bounce. The economic approach gives little insight into which formulae to use because the indices are all exact under a certain set of preferences, and therefore no clear conclusions can be drawn.

41. Assessing the indices methods against the statistical approach provides evidence to suggest that no single index is suitable to cover the full range of items in the consumer basket of goods and services. The axiomatic and statistical approaches to index number theory identify different indices as being more appropriate for different categories of items. (Mayhew, 2018)

42. When considering the theoretical properties of price index methods, the effect of operational choices is also worth looking at, as they may result in a reduction of the differences between methods. Certain theoretical properties, such as additivity, are not truly maintained if an extension method is applied, therefore the satisfaction of the property may not be as useful when comparing methods.

43. **Transitivity**
    Price indices possess the transitive property when all data available across the interval are used simultaneously to calculate the monthly indices. Transitivity can be categorised into both the accuracy and reliability dimension, and the coherence dimension set out in the ONS guidelines for measuring statistical quality. Transitivity is a desirable property for price comparisons because the results will be independent of the choice of base period.

44. Typically, weighted bilateral methods are not transitive, and could lead to production of a drifting index series. By calculating direct indices, rather than chain linking, drifting indices are prevented. However, many studies have shown that chained indices scarcely vary from transitive benchmark indices. (Chessa, 2016a, 2016b, 2017a, 2017b).

45. The Jevons index is transitive for non-changing populations of products, i.e. the sets of products are the same each month. If the product population is changing, transitivity can be upheld by using an imputation method for estimating unobserved prices.

46. The Unit Value index is more efficient when comparing current period to a base period, rather than current to previous, because, the current to previous approach also requires the indices to be chain linked.

47. Multilateral methods offer a solution to the chain drift problem and yield transitive price comparisons. Multilateral price index methods have typically been applied to compare prices across countries or regions but have been adapted to become suitable for price indices across time periods. When a multilateral method is used to generate an index series, each bilateral price comparison depends on prices observed in the other periods of the time window and allows the calculation of transitive indices for each period within that time window.

48. However, transitivity may be lost when index series are updated with data of a new period. The additional periods data might change the parameter values of existing products and create revisions in the price indices up until the current period. Revision of published figures is not allowed in the CPI, apart from exceptional situations. Fortunately, extension methods exist that preserve transitivity without revising.

49. Extension methods that calculate direct indices with respect to a fixed base month produce a drift-free index. To generate indices that are also independent of base month it is suggested to calculate the index of the current month by taking an average of the indices over different base months.

50. The behaviour of index series that are extended using splicing methods is more concerning. Splicing methods are high-frequency chaining methods, which in theory, compromises transitivity. For instance, the window splice method links a year on year index of a transitive series to the index of 12 months ago. That index is a recalculated index and is then different from the published index. This approach of splicing may lead to drift in the published series, since the year on year indices of the published series differ from the calculated transitive series used in the splicing.

51. **Characteristicity**
The Characteristicity criterion takes account of both the accuracy and reliability dimension, and the relevance dimension.

52. Characteristicity is the property that requires transitive multilateral comparisons between member periods to retain the essential features of the binary comparisons that existed between them before transitivity, as prices from distant periods may unduly influence multilateral comparisons. Characteristicity requires that the impact of these influences should be kept to a minimum when they are introduced into the binary comparison. Simply put the multilateral indices between two periods should deviate as little as possible from their binary indices. For example; any multilateral method using a window of nine quarters would have data from over two years previous feeding into its index calculation.

53. Transitivity is particularly useful to circumvent the choice of base period, but a disadvantage is that a transitive index for two period depends on the data of all other periods, which leads

to a loss of characteristicity. The GEKS method can be justified as a means of preserving characteristicity as much as possible.

54. Characteristicity is important to consider when choosing an optimal window length when using a multilateral method; a longer window has higher chance of a loss of characteristicity. According to Ivancic et al. (2011), a 13-month window is probably optimal because it is the shortest window that can deal with seasonal goods with lowest possibility of characteristicity loss.

55. **Flexibility**
It is an advantage for an index method to be flexible enough to be used for a range of purposes, data sources and types of items. The flexibility criterion assesses how methods can be used for new products and data sources and falls into the cohesion and comparability dimension.

56. A flexible index method should be able to facilitate both scanner and web scraped data. Both are high frequency price datasets, but web scraped data is usually absent of weighting information. A flexible method will be adaptable for datasets without weighting information. This is possible by putting equal weight on each product, although this may be considered inappropriate. However, if there are no weights provided by the data then the issue is not whether a method is capable of handling unweighted data, but whether it is appropriate to use unweighted data.

57. A future dataset may also contain more detailed product characteristics. In this circumstance the GEKS and TPD methods can be adapted to produce hedonic indexes; by substituting Time Dummy Hedonic indexes in place of the bilateral link formula in GEKS or using the TPD model in other methods. These hedonic indexes have the potential for better quality adjustment than the basic multilateral methods. There is no recognised approach for adapting the GK method to produce hedonic indexes, although it has been suggested treating each combination of characteristics as a single product identifier as an appropriate approach, although it is unproven whether this approach adjusts for quality change as well as a hedonic index might.

58. Another possible requirement of flexibility is the production of currently unestablished areas of price statistics at ONS, such as spatial indexes. This is feasible for the multilateral price indexes which have been adapted from spatial comparison methods.

59. **Interpretability**
The accessibility and clarity dimension is relevant for this criterion. It is vital for all statistical agencies to be transparent with the statistics they produce, and justify the chosen methods used in published statistics. There are two aspects to the interpretability criterion; firstly, how easily understood the methods are for users to understand and; secondly, whether the price movements that each index produces are easy to interpret, particularly the products/categories of greatest influence and why.

60. Each multilateral method can be explained in quite simple ways, although they are, unsurprisingly, all more complicated than bilateral index methods. Of the multilateral methods the GEKS approach is perhaps the simplest as the multilateral movements are derived by combining superlative bilateral indexes. The GK can be presented using simultaneous equations or matrices. If more detailed characteristics are available a championed approach is to use hedonic modelling, one example of this is the TPD which is

best interpreted using simultaneous equations or matrices, showing the relationship it models between price, product and time.

61. Extension methods also affects interpretability; the direct extension and movement splice methods are easier to comprehend than the window splice methods.

62. The price movements can be investigated using a variety of techniques. A useful analytical technique is the decomposition of the aggregate index movements into a simple sum or product of individual products contributions, allowing the products with the greatest influence on the index movement to be identified, and enabling the justification of the index movements with greater clarity.

63. As with the index methods themselves, bilateral index movements are easier to understand than multilateral movements, and decomposition is simpler as they are expressed in terms of product prices and weights in two periods, whereas multilateral price movements are more complicated to decompose requiring the prices and weights across the chosen multilateral window.

64. Some progress towards the decomposition of multilateral indexes has been made. There are already decomposition methods developed for GEKS and TPD; where the method is broken down into a bilateral Törnqvist term as well as several other factors. (see Van der Grient 2010, de Haan and Hendriks 2013). This method makes it possible to recognise the product groups with the most influence on index movements. The existence of decomposition methods is favourable in terms of greater interpretability for the GEKS and TPD than other multilateral methods. However, it would be a mistake to disregard the GK method because it is an additive method. Therefore, the quantity index it produces has a natural additive decomposition, and decompositions for the associated price indexes are also possible.

65. Perhaps most important is the acknowledgement that interpretability is subjective to the user and audience, and different methods may be preferable for different audiences.

66. **Cohesion**
Unsurprisingly this criterion coincides with ONS cohesion and comparability dimension. The cohesion criterion is twofold; internal cohesion across different areas of the basket, and different data sources and; external cohesion with other NSI to aid comparison between countries.

67. By their theoretical properties no single index method is appropriate for all of items in the consumer prices basket of goods and services. However, to aid cohesion, as few methods as possible should be used. A large proportion of the current UK Consumer Prices Index including owner occupiers' housing costs (CPIH) basket uses the Jevons bilateral method to produce price indices, so it may be sensible to adopt this approach for alternative data sources. However, doing so would mean missing out on the chance to use available expenditure data and no new products could be introduced over the reference period making it not suitable for those categories that experience a high rate of product churn.

68. Lowe is traditionally used above the EA level and so chosen methods should also be cohesive with this method.

69. Jevons is used in EA in current CPI methodology and is included as a “dynamic method” in Eurostat guidance (2017) on processing scanner data. This method is used by many

countries in Europe for supermarket scanner data but can also be applied to web scraped data as it does not require expenditure information.

70. The paper has already discussed the possibility to combine any extension method with any multilateral method. Below are some examples of combinations of multilateral methods with extension methods that are already used by NSI in their national consumer price statistics, or that have been suggested for that purpose:

- GEKS-T in combination with the mean splice method. This combination is currently used in production for the Australian CPI;
- Geary-Khamis method using a FBME (fixed base monthly expanding) window. Chessa (2017b) compared rolling windows with FBME windows. The GK indices were essentially the same for both extension methods; Statistics Netherlands uses the Geary-Khamis method for almost all scanner data. It uses a 13-month window for every type of product;
- FEWS (fixed effect window splice) applies a full window splice to TPD indices;
- RYGEKS stands for “rolling year GEKS”. This method applies a movement splice to GEKS indices.

Conclusion

71. In the introduction an emphasis was placed that this paper is not intended to provide recommendations on methods for price index compilation, but to form ideas about the appropriateness of methods when applying these to alternative data sources.
72. If a certain method is preferred for its properties, it is sensible to use variants of similar methods for different data types and products; this promotes coherence, aids interpretation, and reduces the potential complexity of processing systems.
73. The most appropriate choice of index formula is circumstantial; the best elementary aggregation method will differ depending on which data sources are available to the NSI and the characteristics of the specific data set. The situation in which each NSI operates, and the more broadly used methods in compiling their consumer price indices should also be considered as an important factor.
74. Scanner data should contain timely expenditure information which permits the calculation of weighted bilateral indexes, adjusting for consumer substitution across time. The assessments in Chessa (2016b) and Chessa et al (2017) when comparing weighted and unweighted methods show that differences can be considerable. These results show the importance of choosing a method that uses weights at product level.

75. If you are unable to weight at product level, for example when using web scraped data which has no expenditure information, the method most commonly used across NSI is the use of a chained Jevons. Missing prices can be imputed to attain transitivity. It is not uncommon to see NSI use the Jevons index, with equal weights to avoid chain drift, even if expenditure information is available but using an expenditure filter to deselect low expenditure products impacting the index.

76. However, introducing an expenditure filter can result in products being continually included or excluded, when not necessarily appropriate. For example, if products with low

expenditures are on discount in certain months, the resulting increased sales could raise their expenditure shares to above the threshold and therefore distort the price index.

77. Traditional methods have weaknesses when chaining price indexes at a high frequency (Ivancic, Fox and Diewert 2011). Bilateral methods should be applied with care to data sets that contain expenditures at GTIN level as well as prices. Frequent chaining of weighted bilateral indices become sensitive to chain drift, so when using weighted methods, the direct extension is the preferred approach, as opposed to chain linking. Although drift behaviour can be reduced by using direct methods, no weighted direct and chained bilateral methods are transitive, except for special cases; for example, when expenditure shares are constant for each product over time. However, the direct method will mean new products that are added to the market segment after the base period are left out throughout the year until the index is rebased again, therefore it may not capture the full character of the market segment.

78. One prominent multilateral method is the fixed-effects window-splice (FEWS) index which produces non-revisable quality-adjusted price indices. The FEWS index may be unsuitable for product areas with seasonal variation or high rates of technological advance, where product characteristics change regularly. However, if differences in product quality are slight and product churn is slow, this approach does not improve upon the simpler multilateral methods, such as GEKS, which make no imputation for the changing quality of the sample. Therefore, quality change for that item category must be taken into consideration before choosing a suitable index method.

79. Extension methods are a field of ongoing research across statistical bodies, and chain drift appears to be a prominent issue in recent studies. The calculation of direct indices on monthly extended or rolling windows with respect to a fixed base month yields indices that are free of chain drift. The GEKS-J and the Chained Bilateral Jevons index also performed well and pass all axioms other than price bounce. The GEKS methodology eliminates chain drift. Therefore, if a method that does not suffer from chain drift is desired then the GEKS-J method is recommended for web scraped data and GEKS-T when expenditure information is available. The Törnqvist index is often preferred to Fisher because it is easier to work with in the GEKS from an analytical point of view. The mean splice extension method is suggested by Diewert and Fox (2017), in combination with the GEKS-T method and this combination is currently used in production for the Australian CPI.

80. Time windows with a length of 13 months are chosen as standard to account for a full year of seasonality. The comparisons in Chessa et al. (2017) and Chessa (2017b) shows that the differences between windows of 13 months and 4 years are minor.

81. Table 2 shows a summary of the methods concluded as most appropriate for the use in ONS consumer price statistics, at the elementary aggregate level, when using web scraped or scanner data.

Table 2. A summary of the methods chosen to be taken forward for investigation for use in ONS' consumer price statistics, at the elementary aggregate level.

Method	Weighted	Extension method	Bilateral /Multilateral	Why
Jevons	N	Monthly Chain Link	Bi	Jevons is used in EAs in current CPI methodology and is included as a "dynamic method" in Eurostat guidance (2017) on processing scanner data. This method is used by many countries in Europe for supermarket scanner data (with an expenditure filter), but can also be applied to web scraped data (without expenditure filter) as it does not require expenditure information. The Chained Bilateral Jevons index passed all theoretical property axioms other than price bounce and is transitive (provided missing prices are imputed).
GEKS-J	N	Mean Splice*	Multi	The GEKS method performs well against the theoretical property criterion, preserves characteristicity and can be adapted should more detailed product characteristics become available. It is also one of the most easily interpreted multilateral methods. The use of Jevons as the bilateral indices used to calculate the GEKS-J index means it is suitable for data without expenditure information. This method paired with the mean splice extension method is cohesive with other NSIs.
GEKS-T	Y	Mean Splice*	Multi	GEKS-T is selected for many of the same reasons as GEK-J, however using Törnqvist as the bilateral indices means if expenditure data is available weighting can be done at the product level. The Törnqvist index is preferred to Fisher because it is easier to work with in the GEKS from an analytical point of view. This method paired with the mean splice extension method is cohesive with other NSIs.
RYGEKS-T	Y	Rolling Window	Multi	RYGEKS-T is an alternative option to the GEKS-T with mean splice. The

		(13 months)		use of a rolling window should preserve characteristicity.
Geary-Khamis	Y	Expanding Window	Multi	The GK index is transitive because both the turnover index and the weighted quantity index are transitive. Stats Netherlands use the GK method for every type of product in their CPI. The expanding window extension method means that it is coherent with Eurostat regulations.
FEWS (TPD)	Y	Window Splice	Multi	FEWS is recognised across NSI although only currently used in production by Stats New Zealand for rental prices. It produces non-revisable quality-adjusted price indices.
TPH	Y	*	Multi	Time dummy hedonics index is the ratio of expenditure-share weighted geometric means of quality-adjusted prices. It requires rich data to model effectively, but when the data is available the method is very effective and allows removal of heterogeneity, "noise" or other interfering factors that affect prices and price movements; e.g. seasonality, non-standardised products.

Note: *indicates a method where a final decision is not concluded on most suitable extension method and future ONS work may consider an extension method other than the suggested.

Future work

82. This paper is part of the "index calculations" module of the proposed pipeline for processing alternative data source, which is discussed in APCP-S(19)04 Alternative data sources roadmap. As such the desire of this and future index methods work is to aid ONS decision makers as ONS builds towards using alternative data sources in the production of consumer price statistics. The proposed follow on from this paper is to take the shortlisted methods (found in Table 2) and apply them to the areas of the basket which have been predetermined as suitable for the use of web scraped or scanner data, and recommend which methods are appropriate for each item, these items include; laptops; package holidays; clothing; used cars; and more.

83. It is also in the scope of the follow-on paper to investigate the shortlisted methods behaviour with differing features that may be seen in alternative data sources. The approach will be to apply the appropriate methods to synthetic data sets which have been created with these features; high price variance; high product churn; clearance (dump) prices; level price shifts. The idea is that by investigating these possible features now, an appropriate method will be available should future data sets carry these characteristics.

84. ONS also intends to investigate the impact of product grouping. A downwards bias is strong in items with a high churn rate, for example clothing. The normal pricing pattern for clothing is to come onto the market at a high price as a new item. Then it is discounted, often several times before the item is no longer available when the sale of the item is stopped. At the same time other new items have been introduced to the market at a high initial price, often these items are similar to those that have been previously discontinued. One solution is to group the products and follow the price of the group over time rather than the individual products, thus reducing or eliminating the downward bias picked up. As part of the investigation, the following grouping methods will be analysed for use on alternative data; clustering large datasets into price indices (CLIP); match adjust R squared (MARS); k-means; mean shift; and manual (rules based) clustering. Should ONS decide to implement any of these grouping methods the appropriateness of index methods may need to be revisited.

References

Balk, B. M. (2001). Aggregation Methods in International Comparisons.

Bhardwaj H, Flower T, Lee P and Mayhew M (2017). Research indices using web scraped price data: August 2017 update, Office for National Statistics.

Bethan E (2012). International Comparison of the Formula Effect between the CPI and RPI.

Bentley & Krsinich (2017). Towards a big data CPI for New Zealand.

Bosch O T and Griffioen R (2016). ‘On the use of internet data for the Dutch CPI’. In: UNECE/Eurostat/OECD, Meeting of the Group of Experts on Consumer Price Indices (2016), 2-4 May 2016, Geneva: Switzerland.

Blaudow & Burg (2017). Dynamic pricing as a challenge for Consumer Price Statistics.

Chessa, A.G. (2016a). A New Methodology For Processing Scanner Data in the Dutch CPI.

Chessa, A.G. (2016b). Comparisons of the QU-Method With Other Index Methods for Scanner Data.

Chessa, A.G. (2017a). The QU-method: A New Methodology for Processing Scanner Data.

Chessa, A.G. (2017b). Comparisons of QU-GK Indices for Different Lengths of the Time Window and Updating Methods.

Chessa, A.G., J. Verburg and L. Willenborg (2017). A Comparison of Price Index Methods for Scanner Data.

Chessa, A.G. (2019). A comparison of index extension methods for multilateral methods.

Chris Payne (2017). Analysis of product turnover in web scraped clothing data, and its impact on methods for compiling price indices.

De Hann (2015). Rolling year time dummy indexes and the choice of splicing method.

De Haan, J. and H.A. van der Grient (2011). Eliminating Chain Drift in Price Indices Based On Scanner Data.

De Haan, J. and R. Hendriks (2013). Online Data, Fixed Effects and the Construction of High-Frequency Price Indexes.

De Hann, Willenborg & Chessa (2016). An overview of price index methods for scanner data.

Diewert, W. E., & Fox, K. J. (2017). Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data.

EUROSTAT. (2017). Practical Guide for Processing Supermarket Scanner Data.

Guerreiro, Walzer & Lamboray, STATEC (2018). The use of Supermarket Scanner data in the Luxembourg Consumer Price Index.

Holt, Pham & Webster, ABS (2017). Information Paper: An Implementation Plan to Maximise the Use of Transactions Data in the CPI.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004). Consumer price index manual: Theory and practice.

Ivancic, L., W.E. Diewert and K. J. Fox (2011). Scanner Data, Time Aggregation and the Construction of Price Indices.

Kalisch (2016). Making greater use of transactions data to compile the consumer price index, Australia.

Lamboray C (2017). The Geary Khamis index and the Lehr index: how much do they differ?

Mayhew M (2018). ONS methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data.

Mayhew and Clews (2016). Using Machine Learning Techniques to Clean Web Scraped Price Data Via Cluster Analysis. Survey Methodology Bulletin.

Mehrhoff J, European Commission (2018). Promoting the use of a publicly available scanner data set in price index research and for capacity building.

ONS (2014a). Consumer Prices Indices - Technical Manual.

ONS (2014b). Producer Price Indices - Methods and Guidance.

ONS (2015). Services Producer Price Indices - Methods and Guidance.

ONS (2016). Research indices using web scraped price data: clustering large datasets into price indices (CLIP).

ONS (2017). Research indices using web scraped data: clothing data.

Peter Goodridge (2007). 'Methods explained: Index Numbers'. In: ONS Economic & Labour Market Review | Vol 1 | No 3 .

Rodriguez J & Haraldsen F (2006). The use of scanner data in the Norwegian CPI: The new index for food and non-alcoholic beverages.

Triplett J (2006). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes.

UN Taskforce lead by Chessa (2019). Scanner Data Methods: Instructional Guide. Not yet published.

Van der Grient, H.A. (2010). How to Analyse Changes in Price Indexes when Applying the RYGEKS Formula.

Van der Grient H & de Haan J (2010). The use of supermarket scanner data in the Dutch CPI.

Van Loon & Roels (2018). Integrating big data in the Belgian CPI.

Willenborg (2017). Elementary price indices for internet data.

Willenborg (2017). Transitivizing elementary price indices for internet data using cycle method.

Annex A - Literature Review

Appropriate index methods for alternative data sources

1. ONS have previously highlighted the limitations of using web data to construct price indices, including problems with processing and cleaning large data sets. The research questions whether all web scraped data should be used, potentially reducing the representativeness of the products included in the analysis or whether a sample should be used. (ONS methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data. (2018).)
2. Although bilateral index methods are in use, there is a recognised movement by National Statistical Institutions (NSIs) towards the use of multilateral approaches. This is due in part to these methods, when used with an appropriate extension method, yielding transitive price comparisons, a desirable quality because the results are independent of the choice of base period. (A new methodology for processing scanner data in the Dutch CPI, 2017; Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data, 2017; An overview of price index methods for scanner data, 2017.)
3. GEKS, Geary-Khamis, Time Product Dummy and Hedonics methods reoccur as appropriate multilateral methods alongside suitable extension methods. GEKS can make use of different bilateral indices as inputs but the favoured inputs are Jevons, Fisher or Törnqvist. (An Overview of Price Index Methods for Scanner Data, 2016; ONS methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data, 2018; Making greater use of transactions data to compile the consumer price index, Australia, 2016.)
4. ONS also trialled a new approach called "clustering large datasets into price indices" (CLIP) on web scraped data. CLIP clusters products into similar groups, based on the theory that consumers want to purchase different types of product rather than specific individual products. Clip introduced to ONS the idea of following the average price of a group of products over time, rather than a specific item. (Research indices using web scraped price data: clustering large datasets into price indices (CLIP), 2016)
5. Each multilateral approach can adapt different extension methods to keep the indices transitive as data for the next month is added to the time series. These extension methods can be characterised by one of two method types; a window adjustment method and an index calculation method. Window adjustment methods include a rolling window or a monthly expanding window. Index calculation methods include calculation of a direct index in the current month with respect to a fixed base month, or a splicing method which in turn can be broken down into full window splice, half window splice, movement splice and mean splice. Chessa (2019) concluded that fixed base methods give much better results than window splice and movement splice, acknowledging that fixed base methods are free of drift by definition and finding that window and movement splicing methods show considerable drift, even at retail chain level. Previous studies gave similar results for different splicing methods at COICOP and overall CPI level (Van Loon and Roels, 2018) and a downward drift of window splice was found for lower aggregates (ABS, 2017). Detailed descriptions of these index methods can be found in "A summary of price number methods for alterative data sources" section of this paper. (Making greater use of transactions data to compile the consumer price index, Australia, 2016; Promoting the use of a publicly available

scanner data set in price index research and for capacity building, 2018; Integrating big data in the Belgian CPI, 2018.)

6. Another challenge lies in the choice of window length. Few recommendations have been made on optimal window length, however the most commonly seen window length is 13 months. Research conducted by the Australian Bureau of Statistics (ABS) suggests that if there is sufficient data available then the recommended estimation window should be 25 months or nine quarters. Chessa (2019) also recommends a 25 month window with use of a half splice extension method, but at least one year is recommended for rolling window approaches. Further empirical testing is required to discover optimal window length. (Making greater use of transactions data to compile the consumer price index, Australia, 2016.)

7. Although it is possible to combine any multilateral approach with any extension method, the most commonly suggested combinations are; FEWS (Time Product Dummy method with Window Splice extension), RYGEKS (GEKS method with Rolling Window of length 13 months), mean splice on GEKS-T, and Geary-Khamis with a fixed base monthly expanding window or rolling window.

8. A “December Splice”, an index calculation method using December of the previous year as the fixed base month, is suggested as the preferred extension method over the movement or window splice by the European Commission. This is due to reasons of consistency and interpretability rather than reasons of performance or closeness to a benchmark index. (Promoting the use of a publicly available scanner data set in price index research and for capacity building, 2018.)

9. The ILO/IMF Consumer Price Index manual recommends ‘superlative’ indexes (e.g. Fisher, Törnqvist) as the ideal target for consumer prices. However, the dynamic nature of transactions data can make these methods perform poorly (i.e. traditional price index methods break down when applied to this new data source). (ILO/IMF/OECD/UNECE/Eurostat/The World Bank Consumer price index manual: Theory and practice Geneva, International Labour Office, 2004)

10. The preferred multilateral method is still debated internationally. NSIs tend to consider their choice based on local context. Empirical results typically show little difference between different multilateral methods.

Criteria for measuring strengths and weaknesses of chosen methods

11. The accuracy of a statistic estimated from a sample drawn from a finite population is evaluated by the mean squared error, the sum of its variance and squared bias. This may not be useful for assessing a time series of price and quantity indexes of a scanner or web scraped data set. This is because these data set typically covers the whole population of products; there is no sampling of items required. (An overview of price index methods for scanner data, 2017.)

12. Scanner data can produce quite volatile time series of price indices because of the impact of temporary price reductions (promotional sales), but this volatility should not be smoothed out as it is a genuine characteristic of the phenomenon being measured. Volatility at the product level is less important than bias since we expect volatility to be reduced

considerably as we aggregate across products. (An overview of price index methods for scanner data, 2017.)

13. A theoretical assessment of methods is recommended, allowing for a contrast between bilateral and multilateral approaches and an understanding of the similarities and differences of multilateral methods. (Making greater use of transactions data to compile the consumer price index, Australia. 2016)

14. The ABS produced a framework of considerations whilst implementing new multilateral methods into their consumer prices index from December 2017; this framework is summarised in Table 2. (Making greater use of transactions data to compile the consumer price index, Australia. 2016)

Table 3. Summary of the framework developed by ABS

Criteria	Specific question being answered
Resources	Can we scale up the amount of information used without scaling up manual effort?
Theoretical justification	Axiomatic (test) approach Economic approach
Flexibility	How well can the method make use of datasets with more or less metadata?
Transitivity	Do direct and indirect price comparisons between two periods yield the same result?
Characteristicity	How much is the price change between two periods influenced by prices in other periods?
Interpretability	How easy is it to understand the methods conceptually? How easy is it understand what is driving price movements?

Alternative data sources and index numbers methods currently used by NSI in statistical outputs

15. In countries including Australia, Denmark and Switzerland, supermarket scanner data has been implemented in consumer prices indices using sample-based methods. Typically, the sample design for index methods used are not affected and prices collected in store are simply replaced by prices from scanner data. (Making greater use of transactions data to compile the consumer price index, Australia. 2016)

16. Some NSIs have expressed reservations with using “big data solutions” for scanner data but prefer to stick to a small sample approach. This is due to a lack of international consensus on the best approach and the increase in required resources and consequential workload. (An overview of price index methods for scanner data, 2017.)

17. The Australian consumer prices index previously used scanner data by calculating an average unit value by product, by taking the quantity and expenditure information over the period and using this unit value as a replacement for the directly observed price. Jevons was used at elementary aggregate level and Lowe at the expenditure class. ABS research has resulted in the Australian consumer prices index moving from Jevons to GEKS-T for transactional data.

The ABS recognised that the empirical results from multilateral methods are more sensitive to the choice of extension method but use a mean splice with rolling window of nine quarters. (An overview of price index methods for scanner data, 2017.)

18. Statistics Netherlands are currently implementing a new IT system that aims to use fewer index methods for differing data types and products. This new implementation has led to the use of the Geary Khamis method for all scanner data. (Griffioen and ten Bosch, 2016)

19. Since 2015, Statistics Belgium has produced scanner data indices calculated using the commonly named “dynamic method”, which uses a dynamic basket with a monthly chained Jevons index. The dynamic basket is determined using turnover figures of individual products in two adjacent months, and the product is then included in the sample if it meets a certain threshold. Research results have shown that indices produced by this method do not differ significantly from multilateral methods. The same study recognises GEKS-T as the least biased multilateral method and suggests there is no radical differences between splicing options. For most product groups, web scraping is currently still in the research phase and not used in the calculation of the Belgian consumer prices index. However, in the coming years Statistics Belgium is hoping to include web scraped data in the calculation of its consumer prices index for: Clothing and footwear, Second-hand cars, Student room rental, and Consumer electronics. (Integrating big data in the Belgian CPI, 2018.)

20. Table 4 summarises NSIs’ usage of index number methods when applied to alternative data sources.

Table 4. Some index number methods for alternative data sources used by NSIs

Statistical Office	**Data Type**	**Expenditure data used?**	**Index Method Used**
Statistics Netherlands	Web Scraped (Clothing)	No	Dutot (Chained)
Statistics Netherlands	Scanner (All)	Yes	Geary-Khamis with FBME extension
Statistics Belgium	Web Scraped (Electronics, footwear, hotel reservations, second-hand cars)	No	Hedonics
Statistics Belgium	Scanner (Supermarket)	No	Adapted “dynamic method” using an unweighted chained Jevons index
Federal Statistics Office of Germany	Web Scraped	No	
Australian Bureau of Statistics	Scanner (Groceries)	Yes	GEKS-T with movement splice
Statistics New Zealand	Scanner	Can be weighted or unweighted	FEWS (TPD)
Statistics New Zealand	Scanner (Electronics)	Yes	ITRYGEKS
Statistics New Zealand	Scanner (Used Cars)	No	Hedonics
Statistics New Zealand	Scanner (Electronics, audio visual and	Yes	GEKS-T with movement splice

	household appliances)		
Statistics Norway	Scanner (food, medical, retail, petrol and pharmacy)	Yes	Törnqvist
Statistics Iceland	Scanner (Supermarket)	No	Jevons
Statistics Denmark	Scanner (Supermarket)	No	Jevons
Federal Statistical Office (Switzerland)	Scanner (Supermarket)	No	Jevons

Annex B - A summary of price index number methods

1. A wide variety of price indices have been formulated to measure price changes. These can be grouped into weighted and un-weighted formulae, which can be chosen depending on whether quantities have been observed alongside prices (quantity data are needed to derive expenditure shares, with which price changes are weighted together in weighted indices).
2. Price indices formulas can also be categorised as being either bilateral or multilateral. Bilateral indices only consider two time periods, known as the period of interest, which is the current period and the base period that prices in the current period are being compared back to. By comparison multilateral indices make use of three or more time periods in the index formula.
3. This section presents an overview of the various index number methods that might be considered as suitable candidates for measuring price changes when working with alternative data sources, starting with weighted and unweighted bilateral methods before moving on to multilateral approaches, including multilateral extension methods.

Weighted bilateral methods

4. **Laspeyres (Lowe)**
   The Laspeyres price index can be written as an arithmetic mean of current period price relatives, weighted by base period expenditure shares. The Laspeyres formula (or variants of it) has, and still is, widely used by National Statistics Institutes (NSIs) to construct their price indices at higher aggregation levels, including in the UK. To implement it, a statistical agency needs only to collect information on expenditure shares for the base period, and then collect only price quotes on an ongoing basis. Thus, the Laspeyres price index can be produced on a relatively timely basis without the requirement for having quantity information for the current period.

$$P_L^{0,t} = \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \quad (1.01)$$

   Or alternatively:

$$P_L^{0,t} = \sum_i w_i^0 \left(\frac{p_i^t}{p_i^0}\right) \quad (1.02)$$

   Where:
   $P_L^{0,t}$ = Laspeyres price index between periods 0 and t
   p_i^t = price of item *i* in period t
   p_i^0 = price of item *i* in period 0
   q_i^t = quantity of item *i* in period t
   q_i^0 = quantity of item *i* in period 0
   $w_i^0 = \frac{p_i^0 q_i^0}{\sum_i p_i^0 q_i^0}$ is the base period expenditure
5. It is worth noting that NSI generally would use the Lowe index, often referred to as a "Laspeyres-type index", rather than the above mentioned Laspeyres formula in practice. In a Lowe index the expenditure weights associated with each item are not taken from each indexed period. Usually they are inherited from an earlier period, which is referred to as the

expenditure base period. Generally, the expenditure weights are updated occasionally (although at consistent intervals), to match the base period, but the prices are updated in every period. The Laspeyres index can be considered as a special case of the Lowe when the weight reference period is equal to the base period. The Lowe index is defined as:

$$P_L^{0,t} = \frac{\sum_{i=1}^{n} p_i^t q_i^b}{\sum_{i=1}^{n} p_i^0 q_i^b} \, (1.03)$$

Where:

q_i^b = quantity of item *i* in expenditure base period *b.*

6. **Paasche**

The Paasche price index can be written as a harmonic average of current period price relatives, weighted by current period expenditure shares. The main difference between the Laspeyres and Paasche indices is that, with the Paasche, the expenditure weights are taken from the current period rather than the base period. The lack of information on current period quantities generally prevents statistical agencies from producing Paasche indices on a timely basis, so the Paasche formula is rarely used in the production of national consumer prices.

$$P_P^{0,t} = \frac{\sum_i p_i^t q_i^t}{\sum_i p_i^0 q_i^t} \quad (1.1)$$

7. **Fisher**

There is no justification in favouring taking the expenditure weights from either the base or current periods, hence Fisher proposed taking an average of the two indices to derive a single measure of price change. Taking a geometric mean leads to the so-called *Fisher ideal price index*.

$$P_F^{0,t} = \sqrt{P_L^{0,t} P_P^{0,t}} \, (1.2)$$

8. **Törnqvist**

The Törnqvist index is the geometric mean of the geometric versions of the Laspeyres and Paasche Price Indices. The main feature of the Törnqvist index is that each item’s weight is an arithmetic mean of the expenditure shares in the current and base periods. Therefore, like the Paasche, it tends to be used on historic data sets rather than for the ongoing production of consumer prices, as current period quantity information is needed to calculate the weights. The Törnqvist index has a number of useful features: as well as removing the problem of ‘time reversal’ (if all the data for the two periods are interchanged, then the resulting price index should equal the reciprocal of the original price index.), it can also be used to show each index component’s contribution to aggregate growth because the formula is log-linear, so the contributions can be broken down into additive form.

$$P_T^{0,t} = \prod_i \left(\frac{p_i^t}{p_i^0}\right)^{\frac{(w_i^0 + w_i^t)}{2}} \quad (1.3)$$

9. **Unit value**

   The unit value index is considered quite simple in form. The unit value in each period is calculated by dividing the total expenditure on product i by the corresponding total quantity. In an economic sense; the quantities must be additive, which suggests that they must relate to a single homogeneous item (i.e. the basket must be the same in both periods). The unit value index is then the ratio of unit values in the current period to those in the reference period. It is not a typical price index, as it is essentially a measure of the change in the average price of a single product when that product is sold at different prices to different consumers, perhaps at different times within the same period.

$$P_{UV}^{0,t} = \frac{\overline{P^t}}{\overline{P^0}} = \frac{\frac{\sum_i p_i^t q_i^t}{\sum_i q_i^t}}{\frac{\sum_i p_i^0 q_i^0}{\sum_i q_i^0}} \quad (1.4)$$

10. Prices are typically sampled at a set time/day each month, and each price is assumed to be representative of the average price of that item in that period. However, this assumption isn't always the case, and with this method it is necessary to estimate the unit value for each item. Therefore, data should be collected on both the value of the total sales in a month and the total quantities sold of each specified item in a store. We can then derive a unit value price to be used as the price input into an elementary aggregate formula. It is particularly important to do this if the item is sold at a sale price for part of the period and at the regular price in the rest of the period. Neither the sale price nor the regular price is likely to be representative of the average price at which the item has been sold or the price change between periods. This suggests that the unit value over the whole month should be used instead. With alternative data sources providing more and more data, unit value procedures may be increasingly used. It is worth noting, that the item characteristics must remain constant through time. Changes in the item characteristics could lead to unit value changes that are more a reflection of quantity or quality changes and should not be part of price changes.

Unweighted bilateral methods

11. Unweighted indices are used when quantities aren't observed, this happens at the lowest level of an aggregation structure.

12. **Jevons**

    The Jevons index is calculated as a geometric mean of current period price relatives. The Jevons formula is used in over 60% of items in the UK CPI but cannot be used when prices fall to exactly zero. (Paul Johnson review, UK consumer price statistics - Chapter 10, 2015)

$$P_{Je}^{0,t} = \sqrt[n]{\prod_{i=1}^{n} \frac{p_i^t}{p_i^0}} = \sqrt[n]{\prod_{i=1}^{n} R_i^{0,t}} \quad (2.1)$$

13. **Dutot**

    The Dutot index is calculated as the ratio of average prices in the current and base periods (given a matched basket between both periods) and is usually used for homogeneous items as the formula implicitly gives greatest weight to the highest priced product. The Dutot formula is used for a small number of items in the UK CPI (generally those where the Jevons formula cannot be applied).

- $$P_{Du}^{0,t} = \frac{\frac{1}{n}\sum_i^n p_i^t}{\frac{1}{n}\sum_i^n p_i^0} = \sum_i R_i^{0,t} \frac{p_i^0}{\sum_i p_i^0} \quad (2.2)$$

14. **Carli**
The Carli index is calculated as the arithmetic mean of current period price relatives. The use of the Carli index is effectively prohibited by legally binding regulation for use in the European Union's HICP (which is the UK's CPI), because it can be shown that under certain circumstances the use of the Carli formula, combined with chain linking of in-year indices, introduces an upward bias known as 'chain drift'. Chain drift is caused by the relationship between price change and weight change from the chained short-term overestimating or underestimating the long-term trend of the direct index. Chained indexes are vulnerable to drift when peaks and troughs occur, usually caused by seasonality. If an index is chained at a peak or at a trough, the chained series will be higher or lower relative to a longer term fixed base index.

$$P_{Ca}^{0,t} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_i^t}{p_i^0} = \frac{1}{n}\sum_{i=1}^{n} R_i^{0,t} \quad (2.3)$$

Multilateral methods

15. Multilateral methods have many wanted qualities, both theoretically and in practice, to produce price indexes from big data sources, such as scanner and web scraped data. The theoretical qualities include maintaining transitivity when reweighting and chaining frequently; while, from a practical perspective, automated processes allow a greater sample of products to be used to produce price indexes.

16. Currently no specific multilateral method has received worldwide endorsement. However; the statistics community agree that multilateral methods are the most appropriate approach to produce temporal price indexes when using big data. So far, few countries have implemented scanner data into their consumer prices, selecting to use a variety of different methods. The reluctance of NSI in implementation and the variety of methods of those whom have, is partially due to a lack of consensus amongst researchers about the "best" method to use, and the differences in circumstances for each NSI when producing their consumer price statistics. The availability of data is another issue affecting the implementation of multilateral methods.

17. The variance between the multilateral methods themselves is in the approach taken at the aggregation stage. The Time Product Dummy (TPD) method uses a regression-based approach that estimates price change over time by measuring the statistical relationship between prices, products and time. Whereas the Geary-Khamis (GK) method in general uses a unit value approach to estimate price change and so requires product homogeneity to be able to standardise quantities into common units. The GEKS method applies a geometric mean of the ratios of all bilateral indexes.

18. **GEKS**
The GEKS index was adapted for the time domain by Ivancic et al. [2011]. The GEKS method takes the geometric mean of the ratios of all bilateral indexes (calculated using the same index number formula, for example, Jevons) between several entities. For spatial indexes these entities are generally countries, while for price comparisons across time, the entities are time periods. As web scraped data does not have expenditure information, this article analyses Jevons indices. However, the framework may also be used on scanner data, so we will also analyse the use of Fisher and Törnqvist indices to calculate the bilateral component indices. One problem in using GEKS for official statistics is that when a new time point is added the whole series is revised, to overcome this a NSI would need to use an appropriate extension method.
19. In general, for a GEKS index the formula is:

$$P_{GEKS}^{0,t} = \prod_{l=0}^{T}\left(\frac{P_X^{0,l}}{P_X^{t,l}}\right)^{\frac{1}{T+1}} \quad (3.01)$$

Where X is the chosen bilateral index method.

20. If the formula X used as the input index passes the time reversal axiom, then the GEKS index can be rewritten as:

21. $$P_{GEKS}^{0,t} = \prod_{l=0}^{T}\left(P_X^{0,l} P_X^{l,t}\right)^{\frac{1}{T+1}} \quad (3.02)$$

22. This means that the GEKS-J price index formula is defined as follows:

$$P_{GEKS-J}^{(0,t)} = \prod_{i=0}^{t} ({P_J}^{0,i}\, {P_J}^{i,t})^{\frac{1}{t+1}} \ (3.1)$$

23. The GEKS-T is calculated as the geometric mean of the ratios of all matched-model bilateral indexes $P^{l,t}$and $P^{l,o}$where each period, l, is taken in turn as the base. The GEKS-T method can be expressed as:

$$P_{GEKS-T}^{(0,t)} = \prod_{l=0}^{T} ({P_T}^{0,l} {P_T}^{l,t})^{\frac{1}{T+1}} \ (3.2)$$

24. From the empirical viewpoint it can be expected that the GEKS-Fisher and the GEKS-Törnqvist indices closely approximate each other. The GEKS-F method can be expressed as:

25. $$P_{GEKS-F}^{(0,t)} = \prod_{l=0}^{T} ({P_F}^{0,l} {P_F}^{l,t})^{\frac{1}{T+1}} \ (3.3)$$

26. **Geary-Khamis**
The Geary-Khamis (GK) index was developed for PPPs, but unlike the GEKS, which compares each period to each other, the GK index compares each period to a base period. It is an implicit price index that divides a value index by a weighted quantity index. Using notation like Chessa [2016] it is defined as:

$$P_{GK}^{t} = \frac{\dfrac{\sum_{i \in S^*} p_i^t q_i^t}{\sum_{i \in S^*} p_i^0 q_i^0}}{\dfrac{\sum_{i \in S^*} v_i q_i^t}{\sum_{i \in S^*} v_i q_i^0}} \ (3.4)$$

where the weights w_i are as follows:

$$w_i = \sum_{z \in T} \phi_i^z \frac{p_i^z}{P_{GK}^z}$$

$$\phi_i^z = \frac{q_i^z}{\sum_{s \in T} q_i^s}$$

27. The denominator essentially adjusts for quality, which is why this form of the Geary-Khamis index is also called a quality adjusted Unit Value index. Please note that the Geary-Khamis index appears on both the LHS and RHS of the equation, this means that an exact solution cannot be found, and the price indices are found by solving an eigenvalue problem, or by fixed point iteration. The latter of which is used by Statistics Netherlands.

28. Price indices should only reflect the changes in price of a good or service, not the change in their quality. Quality adjustment is the process by which price indices are adjusted to account for these changes in quality. A hedonic regression can be used to put a value to these changes in quality, by relating the price of an item to its measurable characteristics. There are two main hedonic methods used for Scanner Data, the Time Product Dummy Method and the Time Dummy Method.

29. **Time Product Dummy (TPD)**

The TPD aims to decompose the price of a product into how much of the price comes from being that specific product and how much comes from it being observed in a specific time period. The TPD method uses a regression approach that is like those of hedonic based methods - it uses the statistical relationship between prices, products and time to estimate the decomposition. The TPD model is estimated by pooling together data for a specified window length (T+1) and modelling the log of price against time and product binary indicators. The TPD model is expressed as:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \quad (3.50)$$

where,

$\ln p_i^t$ = log of price for item *i* in period t
α= intercept term
δ^t= time parameter corresponding to time period t
D_i^t= time dummy variable, equal to 1 if the price observation p_i^t was observed to period t and 0 otherwise
γ_i= product parameter corresponding to product *i*
D_i= product dummy variable, equal to 1 if the price observation p_i^tpertains to item *i* and 0 otherwise
ε_i^t= error term

The time effect δ^treflects the overall price level in period t relative to a reference period 0, while the product effect γ_i reflects the typical price of product *i* relative to the reference product N. The intercept α can be interpreted as a general level of prices for a given consumption sector. Transforming equation (3.4), the shadow prices are $\hat{\rho}_i^0 = \exp(\hat{\alpha})\exp(\hat{\gamma}_i)\exp\left(\frac{\sigma^2}{2}\right)$ and $\hat{\rho}_i^t = \exp(\hat{\alpha}^0)\exp(\hat{\gamma}_i)\exp(\hat{\delta}^t)\exp\left(\frac{\sigma^2}{2}\right)$ for all products belonging to an EC. Using the ratio of geometric shadow prices, the price index can be directly estimated from the modelled time effect parameters as follows:

$$P_{TPD}^{0,t} = \frac{\prod_i (\hat{\rho}_i^t)^{\frac{1}{n^t}}}{\prod_i (\hat{\rho}_i^0)^{\frac{1}{n}^0}} = \exp(\hat{\delta}^t) \quad (3.51)$$

$P_{TPD}^{0,t}$ = price movement between periods 0 and t
$\hat{\rho}_i^0$ = shadow price of product *i* from period 0
$\hat{\rho}_i^t$ = shadow price of product *i* from period t

30. A TPD Index can also be defined in terms of the observed prices and the product effects as follows:

$$P_{TPD}^{0,t} = \frac{\prod_i (p_i^t)^{\frac{1}{n}}}{\prod_i (p_i^0)^{\frac{1}{n}}} \exp(\bar{\hat{\gamma}}^0 - \bar{\hat{\gamma}}^t) \text{ (3.52)}$$

31. Here the product parameters can be viewed as adjusting the price for the quality differences of each product so that each price is measured at constant quality.
32. If the shadow prices are estimated using OLS regression then the index is unweighted, if the shadow prices are estimated using WLS with the weights being the expenditure shares, you get a weighted index.

33. **Time Dummy (TPH)**
An extension of the Time Product Dummy model is the Time Dummy Hedonic model. The Time Dummy Model attempts to further breakdown the price of a product into the contribution of the product being observed in a particular period and the contributions for the characteristics of the product. This method is often used in combination with the window splice by NSI for technological goods. Breaking down a price in this way helps to keep the prices observed at a constant quality, as these items often have large quality improvements that frequently take place as models are upgraded, replaced by manufacturers or cease to be available in shops and are replaced with ones with different specifications.
34. The hedonic regression should be based on a large sample to ensure characteristics importance, this regression can be modelled as:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \epsilon_i^t \text{ (3.6)}$$

With item characteristics z_{ik} and time dummy variables D_i^t .
35. The time dummy hedonic index can be written as:

$$P_{TDH}^{0,t} = \exp(\widehat{\delta^t})$$

36. We can think of this time dummy hedonics index as the ratio of expenditure-share weighted geometric means of quality-adjusted prices.

Multilateral extensions

37. When a multilateral method is used to produce a temporal index, each index point depends on prices observed in other periods of the multilateral comparison window. As a result, joining a new period into the multilateral comparison window could alter the index values of previous periods. This creates a revision problem, since published price indexes are only revised in exceptional circumstances. This section describes a selection of methods available to extend a multilateral temporal index when a new period of data is available to form a continuous non-revised index.
38. The choice of length for the estimation and splicing windows has generally defaulted to 13 months (for monthly series) or 5 quarters (for quarterly series), to account for any seasonality in product prices (Ivancic, Fox and Diewert 2011). A paper by de Haan (2015) indicates 25 months or 9 quarters to be a more appropriate window length.

39. **Direct extension**

The direct extension method constructs a price series in a similar approach used to create direct bilateral indexes. When data is available for a new period, the direct extension method recalculates the multilateral comparisons and publishes the index level corresponding to the new period, with the direct multilateral index chained annually (most often in the December month, due to Eurostat HICP regulations); so only the price observations since that last December contribute to each direct index calculation. The direct extension method for a given year can be expressed as:

$$P_D^t = P^l \times P^{l,t} \text{ (4.1)}$$

where,

P_D^t= index level in period *t*
$P^l =$ index level in the latest chaining period *l* before *t* (e.g. the previous December)
$P^{l,t}$= price movement between periods *l* and *t*, using a multilateral window starting in period *l* and ending in period *t.*

40. **Movement splice**

The movement splice method also involves joining a new period into a new multilateral comparison window and extending the index based on a price comparison from this new window. The index level in this new period is calculated by multiplying the previous published index level by the price movement between the previous and the new period, as estimated using the new multilateral window (Ivancic, Fox and Diewert 2011). This is equivalent to the approach used to produce chained bilateral indexes and can be expressed as:

$$P_{MS}^t = P^{t-1} \times P^{t-1,t}(t-T) \text{ (4.2)}$$

where,

P_{MS}^t= index level in period *t*
P^{t-1}= index level in the previous period
$P^{t-1,t}(t-T)$ = price movement between *t-1* and *t* using the latest multilateral window (of length T) to generate a price comparison between *t-1* and *t.*

41. **Window splice**

The window splice method, proposed by Krsinich (2016), uses the rolling window approach to extend the index when a new period of data is available, similarly to the movement splice. However, the methods use price movements from the latest multilateral comparison window to update the index differently. Whereas the movement splice method joins the last period-on-period movement from this window, the window splice method joins on the latest full window onto the index level of *T* periods earlier. This can be expressed as:

$$P_{WS}^{t} = P^{t-1} \times \frac{P^{t-T,t}(t-T)}{P^{t-T,t-1}\,(t-T-1)} \quad (4.3)$$

where,

P_{WS}^{t}= index level in period t
P^{t-1}= index level in the previous period
$P^{t-T,t}(t-T)$= price movement between *t-T* and *t* using the latest multilateral window between *t-T* and *t*
$P^{t-T,t-1}\,(t-T-1)$= price movement between *t-T* and *t-1* using the previous multilateral window between *t-T-1* and *t-1*.

42. **Half splice**

The half splice method is an adjustment of the window splice, splicing on an intermediate movement from the latest window - more than a single period-on-period movement, but less than the full movement. The half splice method can be expressed as:

$$P_{HS}^{t} = P^{t-1} \times \frac{P^{t-\frac{T}{2},t}(t-T)}{P^{t-\frac{T}{2},t-1}\,(t-T-1)} \quad (4.4)$$

where,
P_{HS}^{t} = index level in period t
P^{t-1}= index level in the previous period
$P^{t-\frac{T}{2},t}(t-T)$= price movement between periods $t-\frac{T}{2}$ and t using the latest window from *t-T* to *t*
$P^{t-\frac{T}{2},t-1}\,(t-T-1)$= price movement between periods $t-\frac{T}{2}$ and *t-1* using the window from *t-T-1* to *t-1*.

43. **Fixed base monthly expanding window (FBME)**

Fixed base monthly expanding window (FBME) method has most in common with the currently used method. It uses a time window with a fixed base month, which is shifted each year to the next base month. To include data from a new month the time window is extended with each month. Thus, the window would contain two months in January; three in February; four in March and will ultimately reach the full length of 13 months in December of each year.

44. To ensure price indices are free of chain drift the indices are calculated with respect to the base month with the most recent set of parameter values. FBME can be expressed as:

$$P_{FBME}^{t} = P^{(0,t)} \quad (4.5)$$

where,

P_{FBME}^{t}= Price index in period t
P^{0}= Price index level in the base period

45. **Geometric mean splice**
Diewert and Fox (2017) propose the use of a "mean splice" by taking the geometric mean of all the price indexes that are obtained using every possible link period, given the window length.

46. The ABS have identified some issues with movement and window splicing and highlight the use of mean splice instead; the movement splice can yield downward drift due to disappearing items with unusually low prices whereas the window splice can yield downward drift due to new items entering with unusually high prices. However, the mean splice acts more like a movement splice near the start of the window and more like a window splice near the end; mitigating problems with disappearing and new items.