ADVISORY PANELS ON CONSUMER PRICES – TECHNICAL

Minutes

27 October 2023

Teleconference

10:30 – 12:00

Members in attendance
Mr Grant Fitzner (ONS) (Chair)
Mr Mike Hardie (ONS)
Mr Matt Corder (ONS)
Dr Martin Weale
Professor Paul Smith
Dr Jens Mehrhoff
Professor Ian Crawford
Mr Peter Levell
Professor Rebecca Killick
Professor Bert Balk
Mr Rupert de Vincent-Humphreys

Secretariat
Ms Helen Sands (ONS)
Mr James Wilkins (ONS)

ONS Contributors
Mr Liam Greenhough (ONS)
Ms Laura Christen (ONS)
Mr Ahmet Aydin (ONS)
Dr Mario Spina (ONS)
Mr Dawid Pienaar (ONS)

Apologies
Ms Corinne Becker Vermeulen

1. **Introduction and apologies**

   1.1. Mr Fitzner opened the meeting and passed on apologies from members unable to attend.

   1.2. Mr Fitzner confirmed the position of any outstanding actions.

2. **Clothing classification grouping**

   2.1. Mr Greenhough introduced a paper explaining methods to process web scraped clothing alternative data sources which use a supervised machine learning algorithm to classify web-scaped data into consumption segments, then a product grouping method that tracks prices of similar groups of products over time. Mr Aydin presented results of the classification and product grouping pipelines, and highlighted methods of improvement. The panel were asked for input under four headings:

      2.1.1. Whether precision should be prioritised over recall in the classification task, through implementing a confidence threshold and using an Fβ score.

      2.1.2. Whether the proposed quality adjustment methodology using a hedonic regression is acceptable, specifically, considering words as explanatory variables and running the regression with hundreds of word dummies.

2.1.3. Advise on the over-homogeneity problem for groups with single products.

2.1.4. Whether low scores for homogeneity of price relatives fail the grouping model.

2.2. On prioritisation of the model, panel members agreed with the proposed prioritisation of precision over recall. A panel member highlighted the importance to avoid contaminating the sample used in a category, because the potential benefit of adding a genuine item in a category is small relative to how much would be detracted from the accuracy of the target if an item is incorrectly categorised. A panel member added that if there is no correlation between an item not being classifiable and the price movement, then no information is missed. If there is a correlation, further inspection may be necessary. On the performance of the model, a panel member questioned what the minimum threshold should be.

2.3. On the hedonic regression model proposed, a panel member commented that log price instead of price is typically used in hedonic models. In addition, they highlighted their support for a classification tree approach rather than a regression tree because it is not evident why price should be used as an indicator of what a class looks like as is the case in a regression tree approach. Other panel members agreed with a classification tree type approach because they are easily explainable and repeatable. A panel member highlighted the empirical nature of the problem, and that evidence needs to be provided on the different results of models such as indices to determine what model works best. In response, Mr Greenhough highlighted the theoretical and practical problems of using a classification tree type.

2.4. Should the method use a hedonic regression approach, a panel member signposted the lasso approach because it attempts to fit a small number of classifying variables, to make a good prediction and to get larger groups. On the regression model, a panel member questioned whether there is enough information on the characteristics of clothing items to benefit from this type of regression. Mr Greenhough clarified the information available to demonstrate what was possible, alongside the challenges faced.

2.5. On the problem of over-homogeneity, a panel member asked what the current process is when dealing with many small groups. If manual grouping has been done in response, this may be the best approach to stay with this. A panel member highlighted their concern with small groups due to product churn. The panel member suggested it may be safer when constructing a price index to use average price over a number of items even if there is not an exact month to month match of product. The idea of having a minimum sized group was raised, exactly what the minimum size should be would need statistical experimentation.

2.6. In response to low scores for homogeneity of price relatives, a panel member asked what the previous procedures have been. The panel member

suggested these types of groupings may still have value if they previously existed.

2.7. A panel member highlighted price relatives are the correct criterion to look at instead of price levels, and homogeneity of price relative movements should be maximised. Panel members agreed with stabilising relative price changes rather than price levels. A panel member clarified it should be the price change in the hedonic model which matters.

2.8. Mr Greenhough replied that using relative prices may be contradictory to the aim of matching products which emerge and leave the market, and consequently leads to falling indices. Mr Aydin highlighted price levels address product quality issues better than price relatives.

2.9. A panel member supported ensuring that long-run price movements look correct due to previous issues of downward bias in clothing indices. They asked that the ONS undertake some analysis of clothing seasonality and price discounting.

2.10. In support of relative prices being considered, a panel member raised that issues of correctness in classification would then not be a problem. The panel discussed the clothing market and product cycles, to determine what price dynamics should be captured within the index due to the aim of matching products leaving and emerging the market. Mr Fitzner stated the need to consider the implications of applying relative price on real examples.

3. Grocery data cleaning

3.1. Mr Greenhough on behalf of Dr Spina outlined a paper which built upon APCP-T(23)08 presented at July's APCP-T. In the paper presented by Mr Greenhough, different methods of cleaning grocery scanner data were introduced, and the relevant impact the methods have on the final indices were presented. The panel were invited to comment under three headings:

3.1.1. Whether the proposed outlier detection strategy should specifically target dump prices and remove those observation from index calculation or not.

3.1.2. Provide thoughts on the recommendation to use a price filer with fence of [0.25,4] in combination with a price-quantity dump filter with thresholds $p<05$, $q<0.1$ in outlier detection for grocery scanner data.

3.1.3. General feedback on the analytical results presented in the paper.

3.2. A panel member asked if winsorisation had been considered because it is used elsewhere and it reduces the standard error of the mean. This method reduces the effect of some extreme variables from the sample and reduces RMSE, even if it increases bias. The panel member stated reducing errors rather than reducing bias should be the focus. In reply to winsorisation, Mr Greenhough stated that it may cause clearance products to be included and cause downward bias on the index. Mr Greenhough highlighted that clearance products should be excluded as very few customers benefit from

the reduced clearance product price. Mr Fitzner suggested that more analysis on the price effects of windsorisation be brought to a future panel.

3.3. A panel member wanted to know the reasons for the treatment of outliers. The panel member suggested that there were four possible implicit reasons suggested in the paper, but that these should be made explicit. In addition, a panel member wanted to know what was being filtered out in data cleaning because elsewhere it is considered less of an issue within GEKS-Törnqvist. Mr Greenhough referenced international guidance that indicates GEKS-Törnqvist is still sensitive to dump prices. Mr Fitzner suggested details on the products included in the dump price basket should be provided to the panel, especially if a distinction can be drawn by product perishability vs non-perishability.

3.4. A panel member suggested data that is excluded because of data cleaning may be consistent with rational consumer behaviour. The panel member gave the example of a large business closing down, which led to genuine low prices in the market. In response, a panel member emphasised the dump prices may not be representative and are overweighted in the sample, in which case winsorising the data from the sample makes sense. Mr Fitzner discussed rational consumer behaviour dynamics, and the example of delaying consumption until periods of sales for certain products, where consumers do benefit from clearance prices.

3.5. A panel member wanted to know the purpose of data cleaning. In addition, panel members asked for clarity on the benchmark index, and the true value for which the presented analysis on the results of data cleaning aims. Mr Greenhough presented an example in the context of GEKS-Törnqvist method, which demonstrated why data cleaning is important given clearance prices and how they can downwardly bias the index for two reasons, firstly due to quality issues of clearance product prices, and secondly because only a handful of consumers experience the benefit of dump prices.

4. Publication status of papers

4.1. The clothing classification grouping paper will be published alongside the minutes. The grocery data cleaning paper will be published at a later date.

5. AOB and date of next meeting

5.1. The next meeting will be held on Friday 19 January 2024. Panel members are asked to provide feedback on the suitability of this date.

5.2. Panel members attended a joint meeting with APCP-S to cover impact analysis and a readiness assessment of private rental prices and second-hand cars analysis. This discussion has been redacted.

No.	Action	Person Responsible
1	Panel members to provide feedback on the suitability of the next proposed APCP-T date.	APCP-T Panel members

2	ONS to include details on the types of products typically included in the dump price basket to 1 December publication of Grocery Data Cleaning paper.	ONS staff
3	ONS to bring analysis of the impact of clothing seasonality and discounting to a future panel (date to be agreed against other priorities).	ONS staff
4	ONS to bring analysis on the effects of winterisation to a future panel (date to be agreed against other priorities).	ONS staff