The April APCP-T meeting was held over two dates due to timing requirements for the contingency plan relating to Covid-19. Item 2 was discussed on 7th and 17th April, and items 3-6 were discussed on 17th April.

<div align="center">

ADVISORY PANELS ON CONSUMER PRICES – TECHNICAL

Minutes

Part 1:

7 April 2020

Teleconference

10:30 – 13:00

Part 2:

17 April 2020

Teleconference

10:30 – 13:00

</div>

Members in attendance

Mr John Astin
Prof. Bert Balk*
Dr Antonio Chessa**
Dr Gareth Clews (Methodology, ONS)
Prof. Ian Crawford
Mr Grant Fitzner (Chair – ONS)*
Mr Mike Hardie (ONS)[1]
Mr Peter Levell
Dr Jens Mehrhoff
Prof. Paul Smith

*not present for meeting on 7th April

**not present for meeting on 17th April

Mr Rupert de Vincent-Humphreys
Dr Martin Weale

Presenters
Hazel Martindale*

Observers
Helen Sands*

Secretariat
Ms Joanna Corless (ONS)
Mr Chris Payne (ONS)

1. **Introduction and apologies**
    1.1. Mr Mike Hardie chaired the meeting on 7th April in the absence of Mr Grant Fitzner. Mr Fitzner chaired the meeting on 17th April.
    1.2. The status of actions from the previous meeting were as follows:
        - The stakeholder panel would be consulted in their next meeting on how to better integrate the technical and stakeholder panels
        - Actions 4 and 5 relating to work on the Household Cost Indices were carried forward
        - The remaining actions were complete

---

[1] Mike Hardie chaired the meeting on 7th April

2. Contingency plan for ongoing price collection and index compilation (7th April)

APCP-T(20)06 – *For future publication. Published plans can be found on the ONS website*

2.1. Mr Payne gave an overview of the paper, which detailed plans for maintaining ONS' suite of consumer price statistics during the CoVID-19 pandemic. Panel members were invited to advise on the suitability of the proposed plans for collecting price data and compiling and publishing price indices, and in particular to advise on the most suitable form of imputation for which the market is temporarily unavailable to consumers and on the most suitable treatment for imputing seasonal items.

2.2. The following comments were made around the possible ways of handling items which are currently unavailable to consumers to purchase:

- Members emphasised that the primary objective should be to measure inflation in a valid way. One member suggested that in cases where prices are collected for less than 20% of the usual sample it may still be better to use the available price data than to impute using a parent index which does not represent the item at hand. Another suggestion was that the threshold for using methods such as imputation may ideally vary for different items, depending on factors such as the usual sample size. Another suggestion was that, for items which are temporarily out of stock, we should collect the price, contrary to usual procedures

- There was a suggestion that the correlation between prices in local shops and multiples could be investigated to see whether prices from multiples could be used as a proxy for imputation. The Satio-Vartia-Feenstra price index, which accounts for missing goods without using imputation, was also highlighted for possible use as part of a sensitivity analysis.

- One member preferred the option of carrying forward prices, while acknowledging that there will be some distortion to the timings of price changes as a result. Another disagreed with carrying forward prices and expressed that this doesn't reflect reality in sectors where no goods are available

- One panel member raised concerns about using the all-items rate for imputation, as this will be affected by categories behaving abnormally, such as energy components, and suggested using an appropriate sub-index for imputation. There was acknowledgement that none of the solutions were ideal, however. ONS's approach for items that cannot be purchased was designed to have a minimal impact on the all-items index, which imputing from a sub-index index would not achieve.

- Members raised additional points including: substitution that may be happening due to limited stock should be considered; rather than flagging individual items with limited data, a caveat could be added to the publication as a whole; dropping an item is equivalent to imputing an all-items index

- Consideration should be given to future issues, such as how to deal with the next linking period being affected by current issues, and how to transition from the period affected by the pandemic to the period following it. It will be important to measure price change between the pre-epidemic and post-epidemic period in a way that avoids bias

2.3. Feedback on the tests proposed for identifying imputation methods included:

- Look at the time series statistical property of integration
- Consider effects of the calendar year such as national holiday dates
- The Dicky-Fuller test and time dummy regression are not valid for assessing volatility
- Correlations should be performed on growth rates rather than price levels

2.4. There was a view that there should not be an obligation to follow the HICP guidelines in extraordinary times such as these; in particular, it was argued that NSIs should be able to change weights during the year.

2.5. It was commented that Eurostat regulations do not have to be observed for the compilation of CPIH and RPI.

Action 1: ONS to circulate Eurostat guidance on the compilation of HICP in the context of Covid-19

2.6. There were mixed views about how to address items exhibiting seasonality. One view was that imputing seasonality was important to avoid distortion of annual growth rates, while another was that imputing seasonality should not be attempted as seasonal patterns could be affected for a long period. Another view was that imputing for seasonality may be advisable where seasonal patterns are clear and regular, but not otherwise. One member noted that for some seasonal items, ONS calculates annual growth rates on a year-on-year basis, in which case no seasonal pattern needs to be imputed. A further option for treatment of seasonal items was also suggested: this involved linking the monthly inflation rate for a sub-index from 12 months prior onto the previous month's index.

2.7. Dr Chessa described imputation methods analysed by Statistics Netherlands for air fares and package holidays. The approaches under consideration were i) imputing using a year-on-year index for all items, which would not affect the overall annual rate, ii) imputing using the month-on-month index of the previous year for seasonal items, and iii) imputing using the month-on-month index of a set of reliable aggregates of the current year. Dr Chessa favoured using the annual rate to avoid linking on indices produced in periods just before and during the crisis, which would affect the whole year.

2.8. Around price collection, there was a suggestion that for future contingency planning, retailers could be surveyed in advance to find out whether they have prices available online or would be willing to provide prices by phone.

2.9. Mr Payne noted that publishing a supplementary measure with a rescaled basket, including only those items which are currently available to consumers. There were mixed views about whether such an index should be supplementary or replace the existing indices. It was suggested that ONS could consult the Office for Statistics Regulation over the feasibility of not producing the usual inflation measures for a period of time.

2.10. Some panel members advised considering making revisions when better methods for dealing with certain problems have been developed. Another suggestion was to collect information on weekly expenditures, which could be used for retrospective analysis.

2.11. Other comments included:
- Caution should be taken over changing treatment of outliers, as there may be unusual price movements due to the unusual circumstances
- ONS should maintain transparency around the methods used
- Weights derived from 2019 data, rather than 2020 data, could be used for compilation of indices in 2021

2.12. Mr Hardie advised the panel that the relevant paper will be updated and recirculated to panel members.

Action 2: ONS to update Covid-19 contingency plan paper and recirculate to panel members

3. Next steps for the CPIH historical series (1947 to 1987)

APCP-T(20)07 – This is work in progress and will be published in the future when additional analysis has been undertaken

3.1. Mr Payne introduced the paper, which described the proposed approach to be taken for a CPIH historical series from 1947 to 1987 based on the current modelled CPI historical series over the same period. This follows revision of CPI data between 1988 to 1996 which affects the earlier CPI estimates. The intention was to publish the completed back series by the end of 2020, although this would depend on other priorities.

3.2. Panel members requested to be able to see the model used to generate the series. It was raised that there were some unusual cyclical patterns which could be flawed. It was also suggested that all details of the model used to generate the series should be published to provide transparency for users.

3.3. Panel members expressed that while there was no perfect solution, it was useful to have a consistent series for people to refer to. It was clarified that the published 1947-87 data, as with CPI data over the same period, would not be classified as an Official Statistic and its limitations would be clearly explained.

3.4. The possible uses of the CPIH historical data were discussed. These included informing monetary policy and academia.

3.5. A question was put to the panel around whether it was preferable to reuse the existing CPI data for modelling the CPI back series or whether to remodel it using a similar approach, but incorporating the recent revisions to the 1988-1996 modelled series to produce a consistent set of CPI and CPIH estimates. The panel preferred the latter.

Action 3: ONS to share with the panel the model used to construct the CPI historical series.

4. Classification

APCP-T(20)08 - This is work in progress and will be published in the future when additional analysis has been undertaken

4.1. Dr Martindale introduced the paper, which detailed progress to date on developing automated methods to classify web-scraped clothing data, including: refining consumption segments; sampling and manually labelling a set of web-scraped data; semi-supervised learning methods to automatically expand the number of human-labelled data; and supervised methods to produce classification models.

4.2. It was suggested that homogeneity should be more precisely defined. It was emphasised that homogeneity referred to the similarity of products, and price distributions were not necessarily an appropriate way to assess this.

4.3. The panel raised that research by Eurostat using supermarket data, previously shared with ONS, and supported by a recent hackathon on microprice data organised by the European Central Bank, showed that the way in which features were generated and the percentage of labelled data were more important than the algorithm used to classify data. The hackathon showed that comparatively simple methods performed best; specifically the bag-of-words model and logistic regression.

4.4. It was suggested that 20,000 labelled data points may be enough to train the classifier. The following strategy was put forward to extend the labelled dataset: label 1% of data to train model, then use model to predict labels, then do manual cross check of at least 10-15% of

labelled data. Previously Eurostat checked 10-15% labelled data for supermarkets, but more may be needed for clothing.

4.5. There was discussion around plans for training the model going forwards. It was clarified that twelve months of data would be labelled to cover items from all seasons; in addition, new data would be labelled every month to account for new types of products appearing in the sample.

4.6. On assessing performance of the classifier, suggestions included:
- Looking at the impact of different classification algorithms on prices indices
- Looking at the stability of classification over time
- Looking at the impact on the overall index
- Manually double-checking the classification of the most important items
- Look at not only the favoured class items are put into but also the next-most favoured classes

4.7. It was noted that clothing may be particularly difficult to classify and enquired what plans there were for classifying other types of goods. Work had begun with web scraped technological goods and methods for classifying scanner data would be looked at; the latter may be aided by retailer classification and product codes, and may or may not require machine learning

4.8. There was discussion around international work on the problem of classification. The UN High-Level Group for the Modernisation of Official Statistics (UNECE HLG-MOS) has created a [Machine Learning Project](#) and the UN Global Working Group (GWG) on Big Data will be looking into classification soon.

4.9. Other comments and suggestions included:
- With reference to gradient boosting and neural networks, the more non-linear the method becomes, the more training data you need; 1% would not be sufficient
- ONS could look into reinforcement learning, which may take a unit value bias cost function into account and produce better indices
- A potential issue with labelling propagation and PU learning is that it is self reinforcing, as the output of the learning is taken as the ground truth for the classifier, but may be incorrect
- There is a trade-off between homogeneity and performance and it may be desirable to formalise the given cost function in terms of the homogeneity loss. If willing to accept unit value bias as a loss measure for homogeneity, alternative learning methods could be used.
- Product codes be used as a feature for training the classifier, as they map to particular classes of goods.
- The hierarchy on websites could be used to pre-classify data.
- Algorithm performance with real data could be compared to performance with scrambled data to test whether the algorithm is actually learning.

4.10. The panel should receive an update on further progress with classification work in October 2020.

4.11. Mr Fitzner suggested that a seminar on classification be held at the Data Science Campus

Action 4: ONS to organise Data Science Campus seminar on classification with machine learning.

5. Review of standard errors in CPI

APCP-T(20)09 - This is a draft for future publication

5.1. Prof. Smith introduced the paper, which was a review of published work on estimating errors for a Consumer Price Index. This follows from an action in the September 2018 APCP-T meeting to publish work by Jim O'Donoghue on this subject as an academic paper. The original analysis and the current literature review are planned to be published separately.

5.2. The paper was well received by the panel, who found it to be very useful.

5.3. There was a view that the different sources of error should be reported separately, at least initially, while a more comprehensive measure could be considered later depending on user need.

5.4. There was a suggestion that a metanalysis could be performed with data across various countries.

5.5. There were different views about the best approach to take. One panel member expressed a preference for the virtual replication method for estimating variance. This involved splitting a sample into two parts, rather than running two separate surveys. Another panel member recommended the bootstrap method, which involved resampling of the clusters within the strata with replacement from the available cluster.

5.6. A third panel member raised the question of the extent to which procedures which assume a purely probabilistic design can be applied to scenarios which use non-probabilistic methods, such as internet surveys. On this basis, the panel member preferred a model-based approach, noting that this may treat the selectivity bias introduced by non-probabilistic sampling. In addition, it was acknowledged that most CPIs involve some non-probabilistic sampling.

5.7. It was enquired whether the methods presented in the paper could be used to measure the error in CPI during the time period affected by the Covid-19 pandemic. Mr Payne believed this wouldn't be feasible in the short-term, but could possibly be done retrospectively.

6. AOB

6.1. Mr Fitzner noted the substantial challenge the Covid-19 pandemic posed to the production of price statistics and thanked the panel for their contributions.

6.2. The RPI consultation had been extended to August 21st. An email is to be circulated with the full information on this matter.

Action 5: ONS to circulate update on the RPI consultation to Panel members

6.3. It was suggested that Ian Diamond might be invited to a future APCP-T meeting in person.

Action 6: Ian Diamond to be invited to a future APCP-T meeting

6.4. The planned Masterclass and Conference on Measuring Prices and Quantities, which was cancelled, would be rescheduled once circumstances regarding Covid-19 returned to normality.

6.5. The next APCP-T meeting would take place on 10th July.

No.	Action	Person responsible
1	ONS to circulate Eurostat guidance on the compilation of HICP in the context of Covid-19	ONS

2	ONS to update Covid-19 contingency plan paper and recirculate to panel members	ONS
3	ONS to share with the panel the model used to construct the CPI historical series.	ONS
4	ONS to organise Data Science Campus seminar on classification with machine learning.	ONS
5	ONS to circulate update on the RPI consultation to Panel members	ONS
6	Ian Diamond to be invited to a future APCP-T meeting	ONS
7	Consult Stakeholder Panel over how to better integrate Technical and Stakeholder panels	ONS
8	Redraft the paper on Interest on Financial Debt with the updated expenditure estimates	ONS
9	Provide more information on the Simple Revaluation Approach	ONS