

The January APCP-T meeting was held over two dates to allow further work and discussion of the item on rental indices development plans and progress. Items 3 and 4 were discussed on 15th January and 19th February, and the remaining items were discussed on 15th January.

ADVISORY PANELS ON CONSUMER PRICES – TECHNICAL

Minutes

Part 1:

15th January 2021

Teleconference

10:30 – 12:00

Part 2:

19th February 2021

Teleconference

13.00 – 14.00

Members in attendance

Mr Grant Fitzner (ONS, chair)
Mr Mike Hardie (ONS)
Mr John Astin
Mr Peter Levell
Dr Jens Mehrhoff
Prof. Paul Smith
Dr Martin Weale
Dr Antonio Chessa

Mr Chris Jenkins (ONS)
Ms Natalie Jones (ONS)
Mr Liam Greenhough (ONS)
Ms Jo Corless (ONS)
Ms Helen Sands (ONS)
Mr George Bettsworth (ONS)

Observers

Ms Sofia Poni (ONS)
Ms Tanya Flower (ONS)

Secretariat

Mr Huw Pierce (ONS)
Mr Chris Payne (ONS)

Apologies

Dr Gareth Clews (Methodology, ONS)
Mr Rupert de Vincent-Humphreys
Prof. Bert Balk
Prof. Ian Crawford
Mr Jonathan Athow (ONS)

Presenters

Ms Aimee North (ONS)

1. Introduction and apologies

- 1.1. Mr Fitzner opened the meeting and passed on apologies from members unable to attend.
- 1.2. Two actions were carried forward from the previous meeting. Establishing whether rateable values were used in the CPIH historic series remains ongoing. This will feed into the second

action carried forward: the methodological paper covering the data and modelling used for the CPIH historic series is expected to be published in Q1 2021.

2. Update on price collection

- 2.1. Mr Hardie gave an update on changes to price collection due to the current level of COVID restrictions. Since November 2020, price collection reverted to online only with price collectors working from home. This will also be the case in the base period, January 2021, and will be reviewed on a monthly basis. The list of unavailable items is reviewed monthly and is referred to in the key points of the CPI publication.
- 2.2. An article explaining changes to price collection procedures and the considerations for calculating the 2021 weights, discussed in the [December 2020 APCP-T](#) meeting, is being drafted by Mr Payne for publication on 11th February.

3. Rental Indices Development Plans and Progress (15th January)

- 3.1. Ms North presented on developments to private rental price statistics, covering proposed methodological changes and a review of preliminary results. The new methodology utilises Valuation Office Agency transaction level data allowing for more granular outputs by region, property type and bedroom category. A hedonic regression model is used to predict prices for rental properties based on a selection of property characteristics. Two forms of regression model are under consideration: a general linear model (GLM) with and without interaction terms, and a random forest.
- 3.2. Mr Fitzner highlighted that an [update to the private rental prices development plan](#) would be published on the 19th January 2021.
- 3.3. Dr Chessa noted from his own experience of hedonic modelling the importance of accounting for variable interactions and was pleased to see this was a consideration in the current work.
- 3.4. Dr Weale suggested that some of the anomalous results may be a result of overfitting the model and that removing some of the independent variables might improve performance. He enquired about the functional form of the model, as there were additional considerations for converting a log normal distribution back to a price. Finally, he queried the sample sizes at the most granular level of analysis, as certain local authorities may have very small numbers of certain property types. Ms North acknowledged this concern and described ongoing analysis to determine appropriate combinations of geography and property characteristics that would allow robust estimates.
- 3.5. Dr Mehrhoff highlighted the paradigm change involved in using a machine learning approach (i.e. random forest) for hedonic modelling, asking if it was sufficient for the model to make reliable predictions or if there was also a requirement for the model to be explainable in economic terms. He also highlighted that random forest models perform poorly with time dummies as the fitting process generally discards them, and that this can make the models unstable. He suggested using a shrinkage model (such as lasso) with explicit interactions as a compromise between the control of GLM and the flexibility of random forest. Mr Levell noted that random forest models have an inherent opacity that makes comparisons difficult.

- 3.6. Prof Smith queried the choice of 14 months as the rolling period for the dataset, when 10 months is the average contract length. Ms North explained that 10 months would be ideal, however there are two practical limitations. The first is that relatively few contracts are exactly 10 months long, the average being a balance between six or 12 months which are more common. The second is that data is collected by rent officers who are only prompted to visit properties annually. With a ten-month window this leads to far fewer data points and therefore a more volatile index.
- 3.7. Mr Levell highlighted that changes in the composition of the rental market in recent years makes the hedonic adjustment especially important. The observation that the random forest model gives a different index value to GLM may reveal something about how rental patterns at the local authority level are treated. Comparing GLM with other models that explicitly incorporate local authority interactions could illuminate this question.
- 3.8. Several panel members were interested in more details of the model fitting and asked if a technical paper was in preparation. Mr Jenkins confirmed that there is.

ACTION 1: Ms Jones to provide more detail on the model specifications

ACTION 2: Ms Jones to share a draft of the technical paper with Panel members

4. Update on Private Rental Statistics Development (19th February)

- 4.1. Mr Fitzner emphasised the experimental nature of the results being presented and the rigorous quality assessment they will undergo before being introduced to the headline indices.
- 4.2. Ms North presented an update on progress since the APCP-T meeting on January 15th. Following feedback from that session ONS have tested four further models in addition to the two models presented in January. These are: weighted least squares (WLS) with the interactions that could be implemented and with and without shrinkage applied, WLS using longitude and latitude so that all interactions could be used, and a random forest model with constraints on tree size (“pruning”).
- 4.3. Three of these models have already been discarded by ONS. WLS using longitude and latitude with interactions modelled geographically adjacent local authorities in a similar way, which was felt to be unrealistic. The random forest model using longitude and latitude without pruning applied produced unacceptably volatile results. The WLS model using local authorities with the interactions that were able to be applied and shrinkage applied gave almost identical results to the same model without shrinkage, however it does not provide the statistical metrics (e.g.: R^2 , mean squared error) required to produce observation weights.
- 4.4. The three remaining models for consideration are: WLS with no interactions (hereafter “model 1”), a pruned random forest (“model 2”), and a WLS model with selected interaction terms (“model 3”). The model preferred by ONS is WLS with no interactions, as it is more transparent and easier to quality assure than a random forest and the results from WLS with interactions diverge markedly from the other two.
- 4.5. Mr Fitzner queried an anomaly in the chart of England annual percent growth. Ms North confirmed that the upward trend at the end of the WLS with interactions series (contrasting with the downward trend of the other two models) was most likely explained by the interactions in that model.

- 4.6. Mr de Vincent-Humphreys noted the consistency between models 1 and 2 and asked if this was reflected at smaller geographies and when comparing rental levels instead of index values. Ms North confirmed that all three models were in close agreement at lower geographies, with models 1 and 3 being almost identical. In respect of rental prices, models 1 and 3 were in close agreement while information on model 2 was not available at the meeting.

ACTION 3: Ms North to provide details of the price level estimates for all models at the regional level.

- 4.7. Dr Weale commented that the models predicted a geometric mean of rents (as expected from a logarithmic model) and queried if this was consistent with other ONS price statistics. Ms North confirmed it was consistent with the HPI, had been peer reviewed internally and is consistent with international best practice for housing market indices.
- 4.8. Prof. Crawford asked if confidence intervals could be provided for the models. Ms North confirmed that these were available for the WLS models. Prof Smith recommended bootstrapping as a way of establishing confidence intervals for the random forest model.

ACTION 4: Ms North to provide details of the confidence level estimates for the WLS models.

- 4.9. Prof Smith suggested using cross-validation and analysing prediction errors to give a more objective basis for recommending one model over the others. Prof Smith also raised concerns that the similarity between models 1 and 2 may arise because the pruning mechanism in model 2 has removed all the interactions. Mr Fitzner requested a sensitivity analysis be performed on the models to address these concerns.

ACTION 5: Ms North to perform the sensitivity analysis on the models and share the results with the panel.

- 4.10. Mr Levell asked for greater detail on which interactions were driving the differences between the two WLS models and proposed progressively adding interaction terms to understand the sources of difference. Mr Levell also suggested that controlling for the socioeconomic status of an area (via the ACORN variable) might suppress some of the modelled rent increase.

ACTION 6: Ms North to investigate the use of ACORN in relation to endogeneity

- 4.11. Mr de Vincent-Humphreys enquired about the likelihood of revising the back series of CPI and CPIH in light of these new methods, whether the methods were consistent with HPI in its use of hedonic regression, and for confirmation of the timeline for introducing these new methods into production. Mr Hardie confirmed that further analysis will be carried out during 2021, with a parallel run phase under consideration. No revisions will be made to historic CPI and CPIH values, consistent with ONS policy. Ms North described the parallels and divergences between the HPI and Rents Development processes.
- 4.12. Prof. Balk averred that the opacity of the random forest model was disqualifying, and without understanding the economic meaning of the interaction terms their use was unsafe. This leaves model 1 as the preferred option.
- 4.13. Dr Mehrhoff framed the decision as a balance between the better predictions of a random forest vs the transparency of the WLS model, but only if the random forest's

predictions really were better as measured by mean squared errors or R^2 . The preferred model (model 1) sits well with international recommendations and practices.

5. Weekly online price changes for food and drink items methodology

- 5.1. Ms Sands described the process for compiling the weekly price indices that have been produced during the COVID-19 pandemic, highlighting that this is the first time alternative data sources have been used in a production situation. As many of the decisions taken in the development process were taken at short notice this was an opportunity to review these choices and consider how to take the indices forward. Areas of focus include product identification, outlier detection, and choice of index formula (currently GEKS-J with a five-week movement splice).
- 5.2. Dr Weale queried why the weekly index was using different samples of products to CPI. Ms Sands confirmed that this was partly due to the difficulties in product identification given that local collectors do not collect product IDs, and partly to leverage the greater coverage that the web dataset offers.
- 5.3. Dr Chessa complimented the work of ONS on alternative data sources and highlighted his own research on the topics discussed in the paper, offering to circulate a report on this to panel members. His analysis showed that indices with movement splices are prone to drift and that short windows can create problems with seasonal products. A potential solution could be to use an expanding window instead. Dr Chessa described another approach using a 25-month window with the lack of back data being addressed through imputation, however it was not appropriate for this situation. Dr Mehrhoff observed that 25-months was the emerging international consensus for treating seasonal products. There is a risk however, that if price trends within the basket diverge then the characteristicity of the index may be compromised.

ACTION 7: Dr. Chessa to provide a copy of his multilateral indices research report for circulation to panel members.

- 5.4. Ms Sands described the level of interest from other government departments in the weekly statistics. Mr Fitzner said it was likely that the weekly index will continue in some form after the COVID-19 pandemic due to this level of interest. Further analysis and stakeholder engagement are required to understand how the index can provide the most value in the longer term.
- 5.5. Dr Mehrhoff raised the opportunities that big data provided for identifying emerging trends and enquired about a comparison with conventional headline indices. If the weekly index is to become a permanent publication, then users will require greater detail on its methodology. Ms Sands offered to circulate comparisons with CPIH to panel members.

ACTION 8: Ms Sands to provide comparisons of the weekly index with CPIH.

- 5.6. Mr Astin praised the volume and quality of work by ONS on alternative data sources and enquired if these sources could lead to the end of in-person price collecting. Ms Sands averred that while meaningful volumes of expenditure continued to take place in independent shops there would be a requirement for local collection. Mr Hardie added that the RPI requires local collection, and therefore it would be required to continue until at

least 2030. Beyond that it may be possible to refocus local collection efforts away from supermarkets and more towards independent stores, perhaps in support of regional price statistics.

- 5.7. Prof Smith expressed interest in the distributions of price levels, suggesting that if the distribution was not normal then using a multiple of standard deviation as an outlier detection method was unsuitable. He also questioned the implicit weighting towards stores with larger product ranges. While this might be appropriate for supermarkets vs independent retailers in groceries, in other markets smaller specialist retailers may have larger ranges than larger, more general outlets. Ms Sands confirmed that the price distributions were not normal and improved outlier detection methods were an ongoing topic of research. The choice of product range as a basis for weighting was made due to the difficulties encountered using market share directly. For the retailers being used currently analysis indicates that product range mirrors market share fairly well, however if the number of retailers expands this may need to be revisited.

6. Research indices using scanner data

- 6.1. Ms Corless presented findings from research undertaken using scanner data. Retailers have provided data covering a range of goods, including groceries, homeware, clothing, electronics and fuel. Advice was sought from the panel on three topics:

6.1.1. **Time coverage.** Scanner data is provided by retailers aggregated by week; however, Consumer Price Statistics require prices to be attributable to a single month and some weeks span month boundaries. Proposed solutions are to exclude weeks that straddle months or only use the first three weeks of the month (to align with production timescales), but there is concern that this could introduce bias.

6.1.2. **Product relaunches.** Products are sometimes relaunched with different identification codes creating difficulty in maintaining continuity of data. An approach based on text matching product descriptions is under development, with an intention to test this using a dataset with linked product IDs. However, other NSIs apply alternative approaches such as expenditure monitoring to identify relaunched products.

6.1.3. **Handling discounts:** Discounts can take several forms, requiring a decision on which types to include and exclude from the index. Simple price reductions are straightforward to capture, but multibuy discounts and discriminatory discounts (e.g. loyalty cards) are more challenging. There is also a consideration on whether consistency is required between scanner data and local collection.

- 6.2. On time coverage, Prof Smith suggested constructing a weekly unit value index (UVI) and modelling values for those weeks that are split between months. The weekly index can then be averaged over a calendar month. Alternatively, it may be possible for index construction purposes to redefine a month as a collection of weekly periods. On handling discounts, Prof Smith expressed a preference for capturing as much information as possible, noting that significant proportions of expenditure are covered by loyalty card discounts. One way of addressing this is to treat items bought with and without a loyalty card as separate products.

- 6.3. Mr Astin asked if concerns over introducing bias by discarding weeks could be addressed by testing. Ms Corless agreed that it could. Dr Weale postulated that systematic bias would only arise from discarding weeks if retailers were engaging in behaviour that he found unlikely. While anomalies may occur in individual months, over the long term these were likely to cancel out.
- 6.4. Dr Chessa shared his experience of testing for bias when choosing which weeks to include. For supermarket and fuel data large improvements in accuracy were found when adding second and third weeks to the calculations. In the case of air fares adding a fourth week was necessary for months with moveable holidays (e.g. Easter).
- 6.5. Dr Mehrhoff offered to share the results of his own testing surrounding the time coverage question and encouraged ONS to carry out their own tests. He then described the difficulties that Eurostat encountered using text matching to identify product launches in the absence of other metadata. One approach to mitigating the effect of product relaunches is to use broader product categories. This introduces a risk of unit value bias but give the benefit of greater continuity. For discounts and particularly loyalty cards, Dr Mehrhoff advocated considering time consistency in the index when deciding what to include.

ACTION 9: Dr Mehrhoff to provide details of his scanner data research.

7. Refunds in scanner data

- 7.1. Mr Bettsworth described research into the treatment of refunds in scanner data. Refunds are recorded differently in the datasets provided by different retailers; some log them as individual items while others include them in aggregate statistics. Other NSIs do not generally regard refunds as a problem, especially not for perishable items, although many had not fully considered the issue.
- 7.2. Where refunds are separately logged it is possible to perform analysis on their frequency, and more importantly remove them when obtaining price information. Refunds are far more common in some product categories (e.g.: clothing, jewellery) than others (e.g.: food). Where refunds are included in aggregate figures there are editing techniques that can be applied to extract price information. Current areas of investigation are whether refunds affect the final index and the consequences of choosing one editing method over another.
- 7.3. Mr Astin enquired about the treatment of vouchers, for example discounts offered above an overall spending threshold. Ms Sands confirmed that discounts that were applied to an entire basket rather than a single product are not usefully reported in the datasets. Dr Mehrhoff shared his experience that voucher transactions were recorded as positive quantities and negative expenditures and were likely to be removed by data cleaning processes. Responding to Prof. Smith, Mr Bettsworth confirmed that refunded quantities are reported as negative amounts in the datasets. Prof. Smith asked if this mirrors the treatment of refunds in the Living Costs and Food survey, and how refunds are treated in production currently.
- 7.4. Dr Mehrhoff outlined concerns about timing the processing of refunds. If a purchase is included in an index calculation in one month but refunded after publication, this is difficult to handle in an environment where indices are not revised. The problem is complicated

further when using multilateral indices with splices, and there is no consensus on the appropriate way to handle this. Dr Mehrhoff suggested running case studies on various refund scenarios to understand the possible outcomes, and Dr Weale concurred. Mr Bettsworth highlighted the difficulties of linking refunds to the original sale, noting that other NSIs had also encountered this problem. Mr Levell suggested using information from the retailers that provide disaggregated refund data to model the volume of refunds across all datasets.

- 7.5. Dr Chessa described findings from his own research, confirming that some categories experience a high enough volume of refunds to distort prices. He suggested that if price lists were available it may be possible to build a model that would allow adjustments to be made to the dataset to compensate for refunds.

8. Classification of Alternative Data Sources

- 8.1. Mr Greenhough gave a presentation summarising progress on classifying clothing and groceries in scanner data and outlining further plans for development.
- 8.2. Good progress has been made applying a machine learning approach to clothing, with focus turning to achieving performance gains, understanding the level of bias due to misclassification and preparing the pipeline for production use.
- 8.3. A machine learning approach is thought to be less suitable for groceries however, as products are more likely to have explicit identifiers that can be used instead. Moreover, smaller training datasets and the more granular COICOP structure for groceries increases the risk of misclassification, and the consequences of misclassification for the overall index are more severe for some products due to the higher weight that some groceries carry. Mr Greenhough proposed potentially using machine learning as a method of machine-assistance for improving the efficiency of manual classification, with products still receiving manual scrutiny as assurance. Also potentially exploring whether very low-expenditure products can be excluded without impacting elementary aggregates.
- 8.4. Dr Mehrhoff asked if further information could be provided on the precision/recall performance of the classifiers. A scatter plot of product category weight and precision/recall would indicate if the most important categories for the index were achieving sufficient predictive success. Mr Greenhough confirmed these could be made available.

ACTION 10: Mr Greenhough to provide precision/recall scatter plots.

- 8.5. Dr Mehrhoff noted from his experience firstly that manually reviewing higher weight categories is often valuable, and secondly that classification performance is usually more dependent on the quality of input data than on the choice of classification algorithm.
- 8.6. Prof Smith expressed concern over the risk of propagating incorrect classifications and suggested carrying out further research into the topic.

9. AOB and date of next meeting

- 9.1. Mr Fitzner introduced Ms Poni, who has joined ONS Prices Division as Assistant Deputy Director for development activities
- 9.2. The next meeting will be on Friday 16th April 2021.

No.	Action	Person Responsible
1	Provide more details on the private rentals model specification.	Ms Jones
2	Provide a draft of the private rentals technical paper for circulation to panel members.	Ms Jones
3	Provide details of the price level estimates for all models at the regional level.	Ms North
4	Provide details of the confidence level estimates for the WLS models.	Ms North
5	Perform the sensitivity analysis on the models and share the results with the panel.	Ms North
6	Investigate the use of ACORN in relation to endogeneity	Ms North
7	Provide a copy Dr Chessa's multilateral indices research report for circulation to panel members.	Dr Chessa
8	Provide comparisons of the weekly index with CPIH.	Ms Sands
9	Provide details of Dr Mehrhoff's scanner data research.	Dr Mehrhoff
10	Provide precision/recall scatter plots for the product classifiers.	Mr Greenhough