ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Quality adjustment review

Status: final
Expected publication: alongside minutes

Purpose

1. This paper provides an update on the review of quality adjustment procedures in consumer price statistics that was started last year.

Actions

2. Members of the Technical Panel are invited to:
   a) Discuss the suitability of the proposed quality adjustment indicators, and suggest any other factors that should be included in the system
   b) Advise on the best way of flagging items with unusual quality adjustment indicators for further investigation by ONS staff

Project background

3. The UK Statistics Authority published an independent review of UK consumer prices statistics led by Paul Johnson in January 2015, which made a series of recommendations regarding quality adjustment. The review highlighted that ''ONS should introduce regular monitoring of the impact of quality adjustment on its consumer price statistics'' and how this should include ''monitoring how often non-comparable replacements occur for each item in the basket of goods and services, and investigating those items where this is frequent''.
4. In 2016, a more general review of UK economic statistics was carried out by Professor Charles Bean from the London School of Economics. The Bean Review agreed with the recommendations of the Johnson review and, in addition, proposed having more robust procedures in place to deal with quality change, assessing the suitability of quality adjustment methods more frequently and also collaborating with other NSIs to develop an informed international approach.

Aims and objectives

5. The main objectives of the project are to:
   - Assess the current state of quality adjustment in UK consumer price statistics
   - Make comparisons with other countries
   - Review whether current methods are appropriate
   - Develop and implement a method to monitor the number of non-comparable replacements for each item
   - Specify a process for defining which quality adjustment method is the most appropriate according to the situation

What we do at the moment

6. Following discussions with the Prices Division production team as well as looking through the relevant spreadsheets, it was found that in terms of weights:

- Approximately 54.2% of the CPIH basket is adjusted using Class Mean Imputation (CMI), which is where a new base price is imputed based on the price movement of similar items following the awarding of a non-comparable marker.
- A further 15.8% is adjusted using CMI alongside Quantity Adjustment. Quantity adjustment is used for size or weight changes, whereas all other quality changes are accounted for by CMI.
- 24.6% of the basket is either not adjusted due to the nature of the item not changing in quality (e.g. a passport) or is adjusted using Direct Comparison, which is where we acknowledge a quality change but decide that the change is negligible and treat the replacement as if it is comparable to the original item. Clothing items make up the majority of the direct comparison items.
- The remaining 5.5% is quality adjusted using either hedonic adjustment (0.3%), age adjustment (1.5%), option costing (2.1%), the 'chart collection method' (0.5%) or, in a few special cases, such as mobile phone charges and insurance items, a unique method that is specific to that item (1.1%). An overview of these methods is shown in annexe A.

Quality adjustment in other countries

7. Extensive research been carried out to find out more about the quality adjustment procedures carried out in 12 European countries. This research consisted of E-mail correspondence with contacts from other countries' NSIs and also some online research on the websites of other NSIs. One of the main findings of this research was that the methods used in other countries are broadly similar to those used in the UK, which indicates that our methods are in line with international best practice.
8. Explicit methods used in UK consumer price statistics, such as hedonic adjustment, are only used in a handful of other countries. This is often due to other countries not having the resources available to carry them out. Countries that do not use hedonics include Norway, Denmark, Greece, Finland and Ireland.
9. Although other countries use the same techniques as the UK, there are some differences between the items that they use them for. For example, we use hedonics for laptops, smart phones, tablets and PCs, whereas hedonics is also used for additional technological goods, used cars and residential properties in Germany, and only for clothing and used cars in Sweden.
10. Additionally, it has been found that other countries use methods that aren't currently used within UK consumer price statistics. These include the Expert Judgement approach, used in Germany, Sweden, Austria and Greece, where a team of experts determine whether the quality of an item has changed and the extent to which it has; and the Supported Judgement approach, used in Germany and Spain, which takes into account the energy consumption of an item as well as any follow up servicing and maintenance costs.
11. When enquiries were made regarding the methods that other countries used, many responses were received, however, when asked about the monitoring side of things, respondents were not so forthcoming with replies. We therefore have no evidence that other National Statistical Institutes carry out periodic monitoring of their quality adjustment procedures.

Indicators

Implicit Quality Indices (IQIs)

12. To monitor the impact of current quality adjustment methods on CPIH, ONS has developed Implicit Quality Indices (IQIs), which have been used with the purpose of helping us to identify the item level indices that are being impacted the most by our quality adjustment methods with the intention of flagging those items for investigation.
13. An IQI essentially involves calculating the percentage difference between a Standard Reference Index, an index where no quality adjustment has been applied, and an index which has been quality adjusted. If the value of the IQI = 100 then the adjusted and unadjusted indices are equal to each other. If the value lies between 95-105 then they are within 5% of each other etc.
14. IQIs have been calculated for all food and telecom items from 2013-2017 with some interesting results. Food items generally had values between 95-105 throughout all periods, indicating that there was never more than a 5% difference between the unadjusted and adjusted indices. In each year, it was found that approximately 10 items out of around 170 had values that fell outside of this range, although they were not always the same items. Certain seasonal fruits such as Blueberries and Strawberries fell outside the 5% range in all years, whereas other items only did so in one particular year. It should also be noted that most items that fell outside of this range did so very marginally but some, particularly blueberries and strawberries, had values that fell well outside the range indicating that these items warrant further investigation. What is tending to happen for these items is that an extremely high number of non-comparable markers are being awarded because of a perceived difference in quality between home grown produce in certain seasons of the year and produce being imported from abroad at other times of the year. The results suggest Class Mean Imputation may not be doing a particularly good job on these items at present but it may yet still be the most suitable method as there may be other issues driving the IQI results which we need to consider, e.g. seasonal baskets.
15. The results for the telecom items often fell between 90-110 (10% range) although some items were well outside this range. The reason for this is the diverse range of items in this category. Bundled Telecommunication Services, for example, don't experience a great deal of quality change, whereas other items, such as laptops, do. For the four hedonically adjusted items, there were values as high as 135-140 for smartphones and laptops which were the items that had values lying furthest outside the 10% range. This result was probably to be expected as often when a product of this nature is replaced, the replacement product is often substantially different to the original. The results seem to indicate that using hedonics on these items is justified. The other hedonically adjusted items, namely PCs and Tablets, showed far less disparity, however. The highest value across all periods for Tablets was 118 though values only fell outside the 10% range in 2 out of the 5 years for which testing was carried out. IQI values for PCs never fell outside the 10% range at any time.
16. Another item that produced interesting results was digital cameras, which showed particularly volatile IQI results. Investigation found that the prices of cameras can vary in price by hundreds of pounds, which can lead to a drastically different base price being

imputed compared to the original base price. This item was hedonically adjusted until 2013 and although the decision to stop based on the 2014 Hedonics Review is justified, the retaining of the hedonic sample size of just 16 quotes may not have been appropriate. Therefore, the IQI here may suggest that more prices for digital cameras need to be collected.

17. Some of the more interesting IQI results are outlined in annexe B.

Monitoring of non-comparables

18. As mentioned previously, the Johnson Review recommended that ONS should look into monitoring the number of non-comparable replacements in each item's sample. This is not an easy task as this lack of monitoring appears to be prevalent in other countries too, meaning that there is no standard method for performing this monitoring. ONS has therefore developed three methods for assessing the number of non-comparable replacements in each item's sample.

Method 1: Total number of 'N' Markers as a Proportion of Sample Size

19. The first method is to simply take the total number of 'N' (non-comparable) markers across all time periods within a year and calculate this as a proportion of the total sample size across these same periods. This method is useful in that it gives us an idea of the rate of product churn for an item but does have a slight flaw, however, in that if the number of 'N' markers is very low in the majority of periods but is high in just one then the item may not get flagged as it looks at the overall number of 'N' markers across the year. It is thought that this method could be adapted though to look at the number of 'N' markers in each month instead, which would overcome this problem.

Method 2: The Total Percentage of Imputation in the Sample

20. The second method involves calculating the number of quotes in an item's sample which have been affected by an 'N' marker, or, in other words, all quotes that have an imputed base price. This number is then taken as a proportion of the total sample size across all periods. This method therefore approaches the problem from a different angle to method 1 and has the advantage of taking into account not just quotes with 'N' markers but also those quotes that have been affected by an 'N' marker. It should be noted that when a non-comparable marker is awarded and a new base price imputed, every subsequent period will retain the imputed base price until January the following year when the base price is reset.

Method 3: The Monthly Percentage of Imputation in the Sample

21. The third method is very similar to method 2 but rather than assessing the number of quotes affected by N markers across all periods, this method involves calculating the number of quotes with an imputed base price in each month and finding it as a percentage of that month's sample size. It is therefore the same as method 2 but looks at things on a month by month basis rather than calculating an all year figure.
22. An extract of the results for these three methods is shown in tables 1 and 2 below with the result for method 1 being shown in the 'N Marker Proportion' column, method 2 in the 'Overall Imputed Percentage' column and the method 3 results in the 'Monthly Imputed Percentage' column.

Table 1: Non-Comparable Replacement Monitoring Results for Cheddar Cheese, 2013

Date	N Markers	Sample Size	No. Affected	Method 1: N Marker Prop	Method 2: Overall Imputed Percentage	Method 3: Monthly Imputed Percentage	IQI
Jan	1	116	0	0.55	2.00	0	100
Feb	1	135	0	0.55	2.00	0	100.0017
Mar	1	140	0	0.55	2.00	0	100.0026
Apr	0	143	1	0.55	2.00	0.70	100.0021
May	2	136	2	0.55	2.00	1.47	99.85113
Jun	0	137	2	0.55	2.00	1.46	99.86164
Jul	0	143	3	0.55	2.00	2.10	99.76237
Aug	1	143	4	0.55	2.00	2.80	99.76416
Sep	1	140	4	0.55	2.00	2.86	99.76089
Oct	1	143	5	0.55	2.00	3.50	99.67749
Nov	1	139	5	0.55	2.00	3.60	99.66479
Dec	0	134	7	0.55	2.00	5.22	99.69715

Table 2: Non-Comparable Replacement Monitoring Results for Blueberries, 2015

Date	N Markers	Sample Size	No. Affected	Method 1: N Marker Prop	Method 2: Overall Imputed Percentage	Method 3: Monthly Imputed Percentage	IQI
Jan	22	109	0	20.77	33.20	0	100
Feb	5	134	22	20.77	33.20	16.42	100.0024
Mar	12	128	18	20.77	33.20	14.062	100.005
Apr	32	78	4	20.77	33.20	5.13	100.3012
May	42	78	17	20.77	33.20	21.79	104.0215
Jun	25	81	21	20.77	33.20	25.93	99.40908
Jul	39	100	41	20.77	33.20	41.00	107.4457
Aug	12	108	49	20.77	33.20	45.37	105.115
Sep	9	86	55	20.77	33.20	63.95	105.8334
Oct	33	111	59	20.77	33.20	53.15	106.7782
Nov	20	103	50	20.77	33.20	48.54	105.3902
Dec	8	131	78	20.77	33.20	59.54	116.4602

23. The cheddar cheese results for methods 1, 2 and 3 are fairly standard for food items as the number of non-comparable markers is low (only 0.5% of sample size) and the sample size is relatively stable throughout. The low number of 'N' markers means that only approximately 2% of the overall sample is imputed across all periods with the level of imputation in the

sample gradually rising each month culminating in just over 5% of the sample having imputed base prices by the end of the year.

24. In contrast, the results for Blueberries are an example of an item which has an extremely high number of 'N' markers (20.7% of sample size) which causes great fluctuations in the number of quotes included in the index calculation for the item from one month to another. The results show that approximately a third of the total sample has an imputed base price, which is exceedingly high and the month by month levels of imputation for blueberries are also far higher than one would expect for a food item. Unusually, although the level of imputation generally increases over time, large fluctuations in the sample size each month caused by a combination of 'N' markers, 'M' (missing) markers and 'T' (temporarily unavailable) markers means that there are actually times where the level of imputation falls. Nevertheless, the level of imputation in the sample can be as high as 63%, which is far higher than expected.
25. Using 5% as a cut off point for method 1, approximately 5 food items would be flagged each year based on the testing results of 2013-2017. Adopting a 10% cut off point for method 2, would see approximately 10 food items flagged for investigation each year. If a higher cut off point of 20% for method 3 were to be applied then approximately 10 food items would have been flagged each year. The different cut offs have been used so that only a manageable number of items are flagged for review each year.
26. Using these thresholds, the items being flagged by each method were sometimes the same but sometimes different. Items such as strawberries and blueberries were always picked up by every method and some other items were also picked up by all 3 methods though, unlike the strawberries and blueberries, these were usually only in one year. In general, a handful of items were flagged by methods 2 and 3 only with less items being flagged by method 1 and just one other method. There were also some items that were flagged by only one method.
27. The same three methods were applied on telecom items but because of almost constant advancements being made in technology which results in a far higher level of product churn, the number of non-comparable replacements and the level of imputation in an item's sample is far higher than those of food items. Therefore, different cut-off points have been chosen for telecom items.
28. A further complication is that some items are collected locally by a company ONS has a contract with and others are collected centrally in house. The locally collected items usually have a sample size of anywhere from 100-300, whereas the centrally collected items have sample sizes as low as 15-30. This means that making comparisons between the level of imputation/number of 'N' markers between locally and centrally collected items is quite difficult. To try and account for this, a 10% cut-off limit for method 1, 15% for method 2 and 30% for method 3 have been used for locally collected telecom items; and 15%, 30% and 50% respectively have been used for centrally collected telecom items.
29. Using these thresholds, method 1 flags all of the hedonic items (smartphones, tablets, PCs and laptops) and also cameras in every year. It may occasionally flag a further item or two in a particular year but these additional items don't tend to be the same from one year to the next. Method 2 also tends to flag the hedonic items and cameras but also seems to capture TVs and Blu Ray Players in each year. Method 3 tends to flag more or less the same items as method 2.

Discussion points

Flagging items

30. Although the IQI values are good at showing the impact of quality adjustment on the price index of an item, it is unclear how to go about flagging those items that require investigation. There are 3 possible approaches:
    - The first is to simply use limits such as 5% for food and 10% for telecoms as values that lie outside this range, at least going on the testing results for 2013-2017, are essentially outliers. As mentioned previously, the number of food items flagged by this method would usually end being close to 10 items out of 170 for food and 5 items out of 30 for telecoms. For the period that was tested, this approach seems to do a reasonable enough job but the thresholds used are somewhat arbitrary.
    - A second approach would be to rank each item in terms of how far its most extreme value lies from 100. This would entail identifying the maximum and minimum IQI values for each item during a year and assessing which of the two lies furthest from 100. The difference between that value and 100 is then calculated and used as that item's ranking score. Again though, the cut off point for which items should be flagged is somewhat arbitrary but it is thought that possibly the top 10% for each division of the basket could be used though this is dependent on the number of items in the division and the actual scores. An additional consideration is the standard deviation of each item's IQI values, which may help to indicate whether the most extreme value is a one off anomalous value or if there are multiple values that are higher or lower than expected. An extract of the results for food items for 2013 is outlined in table 3 below.
    - The third approach is to not think about the situation in terms of a threshold but rather in terms of the number of items we can realistically review with the resource we have. This would involve ranking the items like above and then selecting the X highest ranked IQIs.

Table 2: Ranking Scores for Food Items in 2013

Item Description	2013 Max	2013 Min	2013 St Dev	2013 Score
Blueberries, punnet, per kg or per punnet	130.04	98.64	10.59	**30.04**
Soft Continental cheese, per kg	112.81	100.00	3.87	**12.81**
Carton/Box of Chocolates, 150-400g	112.80	99.58	5.83	**12.80**
Strawberries, per kg or per punnet	107.96	98.13	2.60	**7.96**
Chocolate covered ice cream bar 80-130ml	106.22	100.00	2.28	**6.22**
Sugar, granulated, white, per kg	105.86	99.99	2.63	**5.86**
Frozen Prawns, packed, per kg	101.08	94.74	1.89	**5.26**
Continental Deli Type Meat, sliced, 40 - 100g	104.67	99.47	2.24	**4.67**
Fresh Veg, Courgettes, per kg	104.29	99.99	1.38	**4.29**
Peaches/Nectarines, each (SEASONAL)	100.20	95.82	1.38	**4.18**
Frozen Breaded/Battered White Fish, 400-550g	103.85	100.00	1.38	**3.85**

Grapes, per kg	103.83	100.00	1.50	**3.83**
Avocado Pear, each	103.70	100.00	1.31	**3.70**
Plums, state per pack/kg	103.61	96.37	1.59	**3.63**
Frozen Fish Fingers, 8-12 pack	103.35	100.00	1.48	**3.35**
Canned Tuna, specify oil, brine or water,130-200g	100.00	97.19	0.98	**2.81**
Plain Biscuits, 200-300g (e.g. Digestive)	102.81	100.00	0.92	**2.81**
Vegetable Pickle, 280-520g (e.g. Branston)	100.46	97.36	0.88	**2.64**
Small Oranges Type, state per pk/kg	102.43	98.71	1.03	**2.43**
Frozen Beefburgers, pack of 4, specify weight	102.34	100.00	0.86	**2.34**
Dried Fruit, 100 - 250g pack, eg apricots	102.33	100.00	0.95	**2.33**
Frozen Vegetarian Burger/Grills, pack, 200-454g	102.23	99.99	0.67	**2.23**
Pork Pie, individual, approx 3in/8cm	100.58	97.82	0.79	**2.18**
Olive Oil, 500ml - 1 Litre	100.01	97.83	0.93	**2.17**
Fizzy Energy Drink, e.g. Red Bull, Lucozade, 250-500ml	102.13	100.00	0.73	**2.13**

31. These three approaches are all thought to be viable options for the three methods for monitoring the number of non-comparable replacements too.
32. Members of the Technical Panel are invited to discuss which of these approaches they feel may be best for our purposes (both for IQIs and for methods for monitoring the number of non-comparable replacements) or indeed whether they believe that these methods should be applied together. Any alternative suggestions will also be welcomed. Additionally, member's thoughts on whether it would be better to monitor on a division by division basis or doing the whole basket as one would be more viable.

Scoring system

33. There is a certain amount of overlap between the results for the three methods for monitoring the number of non-comparables and the IQI results. In the testing period of 2013-2017, the three non-comparable monitoring methods typically flagged between 10-15 food items between them in a year with approximately 4 of these also being flagged by IQIs using 5% range as a limit. Similarly, there were approximately 10 items flagged by the three methods for telecoms each year with around half of these also being flagged by the IQI results assuming a 10% limit.
34. Due to there being a slight correlation between the results, a scoring system has been developed that takes into account both an item's IQI result, its non-comparable replacement monitoring results and has also been extended to try and take into account the item's weight.
35. The scoring system for food items means that any items with an IQI value outside of the 5% range will obtain a certain score but the size of the score is dependent on how far outside 5% the most extreme value is. If the value is between 5 and 6% then the item will score 1, if greater than 6 but less than 10% then it will score 2 and anything lying outside the 10% range will result in the maximum score of 3. For telecom items, anything outside the 10% range will obtain a score of some sort. If greater than 10% but less than 12% then a score of 1 will be awarded, if greater than 12% but less than 20% then a score of 2 will be given, and anything above 20% will result in the maximum score of 3.

36. The methods for monitoring non-comparables are given a score based on the limits outlined above. If the item is flagged by one of the three methods then it will score 1, if it is flagged by two methods then it will score 2 and if it is flagged by all three methods then in will score the maximum of 3.
37. One final consideration is the actual weight of the item. Clearly the higher the weight of an item then the bigger the effect of any quality adjustment carried out on the price indices above the item level. Ideally, contributions would be used instead of weights but unfortunately these are unavailable at the item level. The inclusion of a weights component has the benefit of ensuring that we have a priority order for items needing investigation and also assists us in deciding the suitability of a method for a particular item, for example, if the item has a relatively small weight then using resource intensive methods on it is probably not worthwhile. How to score the weight of an item is one of the more difficult components for the system but one of the better approaches tested involves adding the IQI and Non-Comparable scores together and then multiplying the result by the item's weight. This has the advantage over alternative options as it means that items are being flagged because of actual quality adjustment and not primarily because of their weight.
38. An extract of the scoring system for 2015 food and telecom items is outlined in tables 4 and 5 below.

Table 4: Scoring System for 2015 Food Items

Item Name	CPIH Weights	IQI Score	Non-Comparable Score	Total Score
Strawberries, per kg or per punnet	0.77	3	3	4.62
Chocolate covered ice cream bar 80-130ml	0.7	3	3	4.2
Bag of sweets, branded chocolate, 100-185g	0.8	2	2	3.2
Blueberries, punnet, state per kg or per punnet	0.42	3	3	2.52
Carton/Box of Chocolates, 150-400g	1.1	0	2	2.2
Frozen Chips, 900g-1kg	0.9	2	0	1.8
Doughnut, each	0.6	3	0	1.8
Packet of Peanuts, 90-200g	0.84	2	0	1.68
Frozen Chicken Breasts, 500g - 1.5kg, per kg	0.48	0	2	0.96
Canned Tuna, specify oil, brine or water,130-200g	0.69	0	1	0.69
Plums, state per pack/kg	0.21	0	3	0.63
Fresh Veg, Courgettes, per kg	0.3	2	0	0.6
Bag of sweets, not chocolate, 150-250g	0.6	1	0	0.6
Cream Crackers, packet, 200-300g	0.48	0	1	0.48
Joint, oven ready, gammon/pork, 450-700g	0.16	0	2	0.32
Pre-Packed Salad, 100-250g, specify type	0.3	0	1	0.3
Dried Fruit, 100 - 250g pack, eg apricots	0.28	0	1	0.28
Small Oranges Type, state per pk/kg	0.21	0	1	0.21
Fresh Turkey Steaks, per kg	0.16	0	1	0.16
Home Killed Liver, per kg, state type	0.16	1	0	0.16
Protein Powder, 900g - 2kg, specify type	0.08	0	1	0.08
Large Loaf, white, unsliced, 800g	0.24	0	0	0

6 Bread Rolls, white/brown	0.48	0	0	0

Table 5: Scoring System for 2015 Telecom Items

Item Name	CPIH Weight	IQI Score	Non-Comparable Score	Total Score
Laptop Computers	1.25	3	3	7.5
Digital Compact Camera	0.72	3	3	4.32
Digital Camcorder	0.58	2	3	2.9
Flat panel TV 33"/82.5cm or larger	0.84	0	3	2.52
Flat panel TV 23-32"/57.5-80cm	0.72	0	2	1.44
Interchangeable Lens Digital Camera	0.7	2	0	1.4
Tablet Computers	1.05	0	1	1.05
Personal Computers (HICP)	0.3	0	3	0.9
MP4 Player	0.21	0	3	0.63
Smart Phone Handsets	0.2	2	0	0.4
Blu-Ray Player	0.06	0	2	0.12
Digital Television Recorder/Receiver	0.06	0	1	0.06
Telephone, not mobile,	0.2	0	0	0
Telephone Charges	1.2	0	0	0

39. In terms of which items need to be flagged for investigation, for both food and telecom items, only a relatively small number of items have obtained a score: approximately 20 for food and 12 for telecoms in each year of testing. These numbers don't seem too high at first glance although testing will need to be carried out on other divisions of the basket to see what the approximate total number of items would be across the whole basket and whether this number is manageable regarding resources. If the total number is too high then some sort of cut-off point may need to be applied.
40. Another factor that we would like to include would be a resource intensiveness score to reflect the additional resource burden of using explicit methods such as hedonics. Though adding a score of some description to the total score for hedonically adjusted items shouldn't be too difficult, it is unclear how great the additional score should be.

Thomas Lewis
Prices, Office for National Statistics
January, 2019

List of Annexes

Annex A	Summary of within-year quality adjustment methods
Annex B	Graphical representation of IQI results

Annex A – Summary of within-year quality adjustment methods

1. The table below gives an overview of each of the quality adjustment methods that are currently used in consumer price statistics and explains when they should be used.

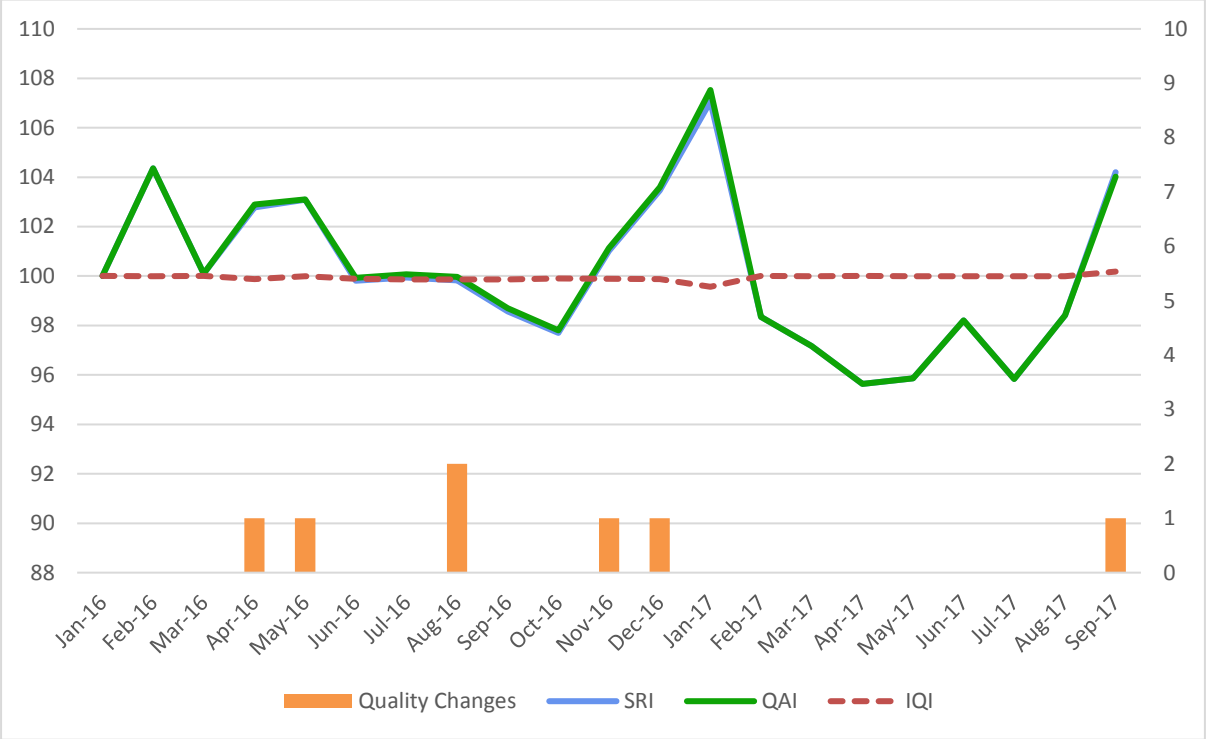
Table A1: Overview of CPI quality adjustment methods

Method	Description	When should it be used?	Where is it used?
Direct Comparison	New product price compared directly to old product price	When the item has been deemed comparable – it provides the consumer with no extra *utility*- e.g. a change in colour.	For majority of CPIH basket when there have been no significant quality changes
Class Mean Imputation / Bridged Overlap	A base price for a replacement product is imputed using the price movements of similar products. The base price that is imputed for ONS price indices is imputed from the price movements at the item level only.	When there are no directly measurable changes. E.g. an orange from UK is replaced with a one imported from Morocco. A base price for the Moroccan orange could be imputed based on the price movement of all oranges.	For majority of CPIH basket when there have been significant quality changes that cannot be accounted for by explicit methods. Often used for items with relatively low product churn, when quality change is significant but not drastic and, to some extent, items with a lower weight in the basket.
Quantity Adjustment	Old product price is pro-rated to make it directly comparable to new product price (e.g. if bigger, the price is increased proportionately).	When there's an observed change in quantity (e.g. packet size, or weight)	For majority of CPIH basket when there have been observed changes in volume or size of products
Option Costing	Change in quality measured by the estimated change in cost of the quality, and the take-up rate of said quality.	When there are simple changes in the physical characteristics of a good. E.g. Parking sensors on a car.	New cars
Monthly chaining and resampling	New samples are selected monthly and chained together using any items	When there is a high product turnover and a	Not currently used – potential to investigate using web-scraped data.

	available in both periods – susceptible to chain drift?	large amount of data to draw samples from	
Hedonic Adjustment	Uses a regression model based on current prices/specifications. Ratio of predicted prices between current and base period used to quantify change in quality.	When there are a combination of measurable characteristics driving the observed price. When there is rapid technology change leading to high rates of product churn.	PCs Laptops Tablets Smart Phones
Age Adjustment	This involves calculating an average of the 3 year and 2 year prices and then weighting them according to the time of year.	This technique is used to account for the price of an item such as a car decreasing significantly with age.	Second Hand Cars and Motorbikes
'Chart Collection Method'	The prices of the top rated/most popular items only are selected and are used to calculate price index.	When the item in question's popularity is timely. E.g. chart album is rated as number 1 one month but due to the release of new albums and the age of the album, its popularity will start to fall.	It is currently used for CD albums, Video Games, Books and Mobile Phone Apps.

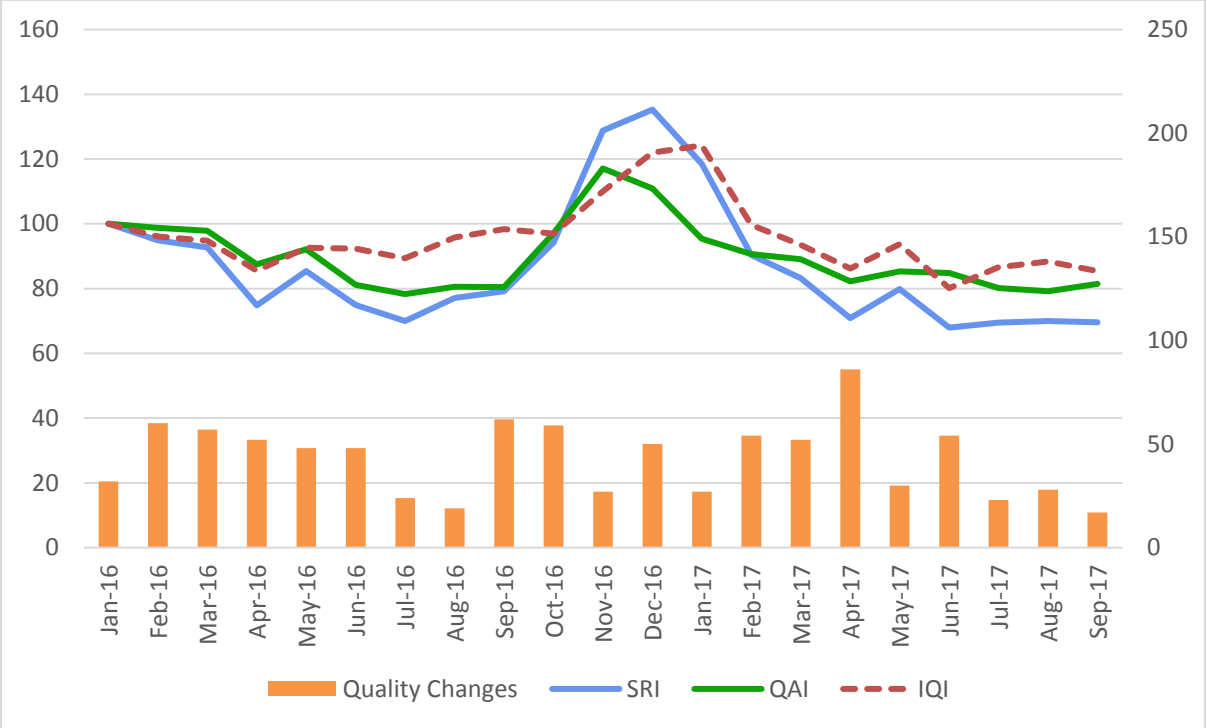
Annex B – Graphical representation of IQI results

Table B1: IQI for spreadable butter



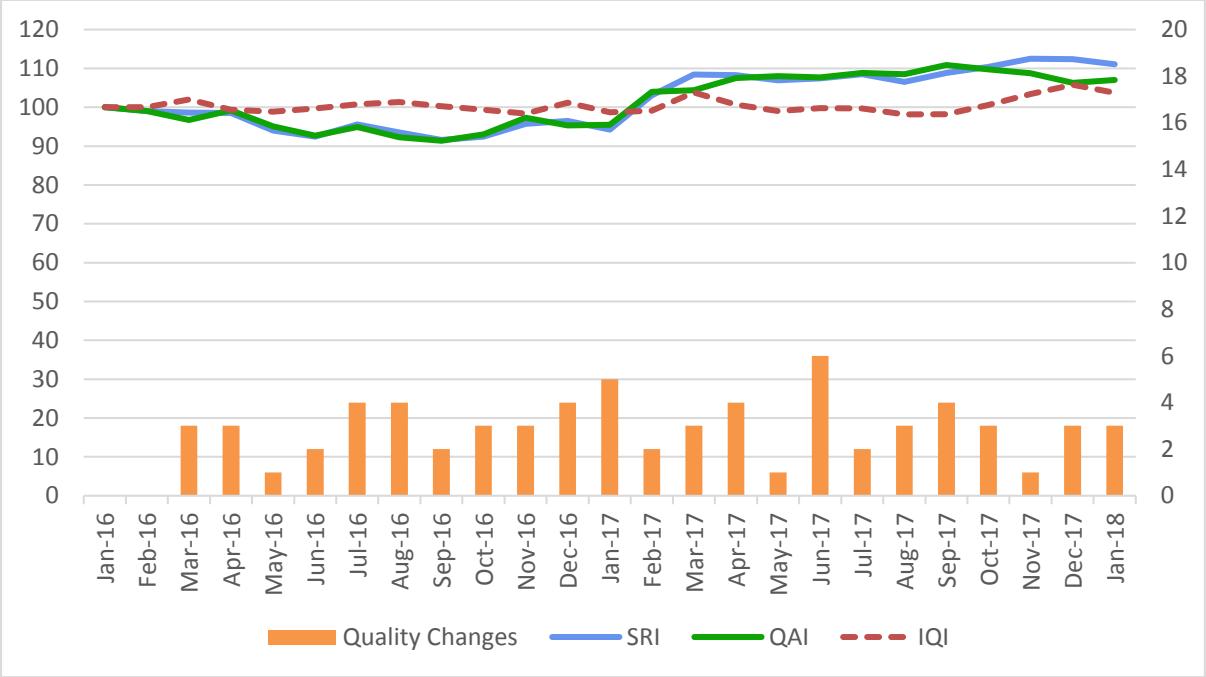
1. The IQI results for Spreadable Butter are fairly typical of food items. The item had only a small handful of noted quality changes: 6 in total. 5 of these were because a replacement item was deemed to be non-comparable and 1 was a weight change. The total number of quality changes is seen in the yellow bars which uses the scale on the right hand side of the graph.
2. Due to the low number of quality changes, the unadjusted index (blue line) and adjusted index (orange line) have very similar values throughout the series to the point where blue line is almost hidden on the graph. As a result of this, the IQI values (grey line) were always very close to 100 throughout.
3. The imputed base prices that were awarded following the non-comparable replacements in November and December 16 had a slight impact on the January 17 figure but this was the only noticeable deviation in the IQI values.

Table B2: IQI for strawberries



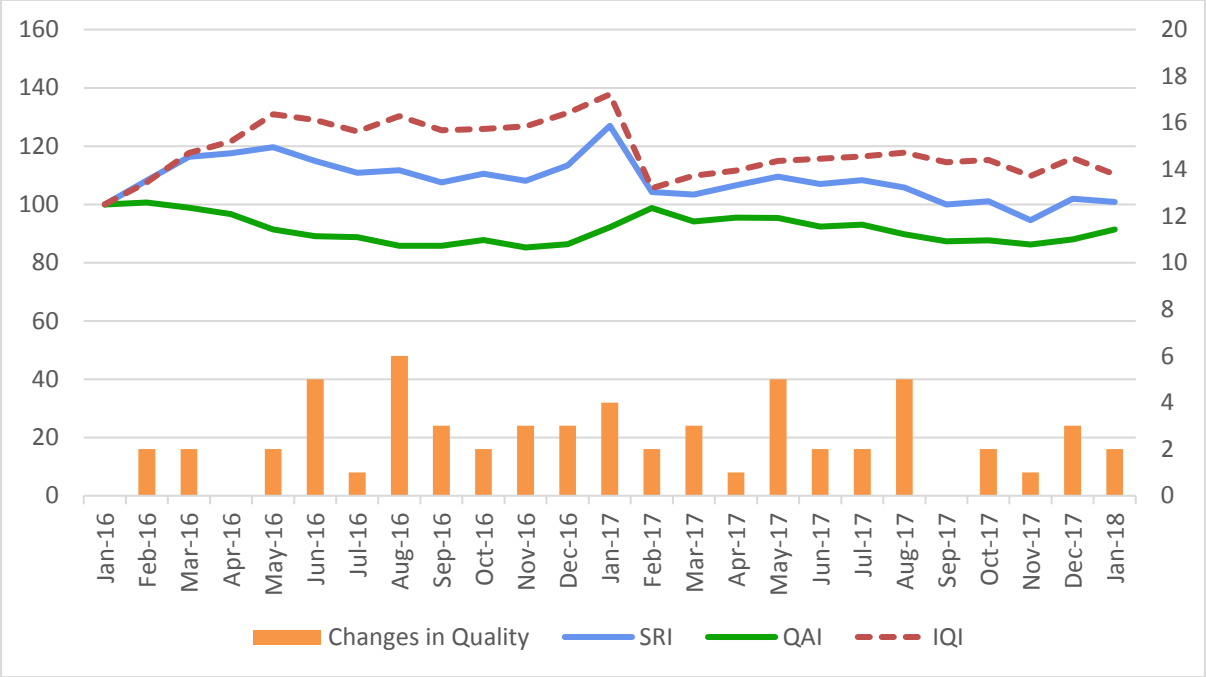
4. The number of noted quality changes for strawberries is clearly far higher than butter across all periods- there were a total of 909: 908 non-comparable markers and 1 single weight change. Yellow bars are noticeably higher than for butter and a different scale was required for the graph.
5. As there are a considerably higher number of non-comparable replacements for this item, a much greater amount of quality adjustment has been carried out. Products with non-comparable markers are withdrawn from the index calculation each month until their new base price has been imputed in the following period. This has a significant effect on the sample size for each month. Additionally, again because of the high number of N markers, those items that are in the sample quite often have an imputed base price that was imputed from an earlier period. In fact, by the end of the year over 80% of quotes for strawberries had an imputed base price.
6. The resultant IQI index (grey line) sees values as low as 80 in June 17 and as high as 124 in Jan 17. This appears to indicate that if our quality adjustment methods are correct, that the quality of a strawberry can change by 44% over the course of just a few months. This result therefore indicates that Class Mean Imputation is not doing a particularly good job at quality adjusting strawberries at present and may not be suitable for items with such a high number of non-comparable markers. That said, Class Mean Imputation may yet still be the best method compared to alternatives but there are probably other issues driving the IQI that we need to consider going forwards e.g. using seasonal baskets.

Table B3: IQI for PCs



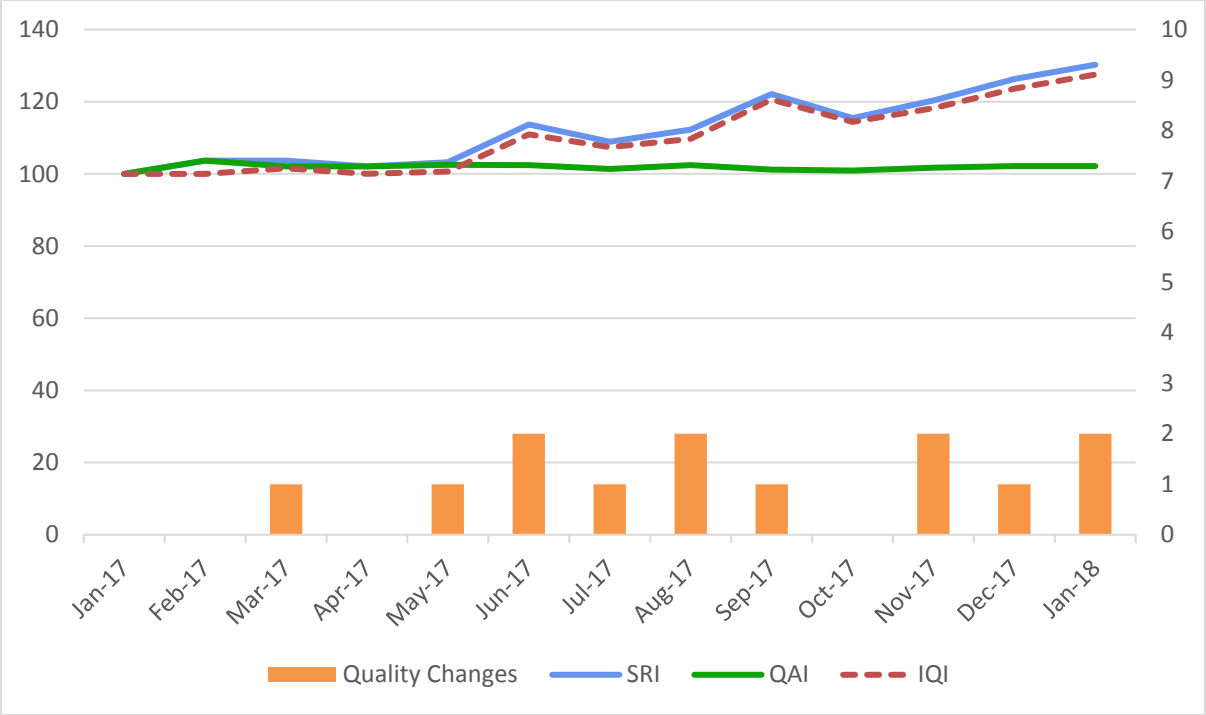
7. PCs are currently quality adjusted using hedonics as it is an item that has a relatively high product churn. The number of noted quality changes is about average for a hedonic item.
8. The Unadjusted index (blue line) and the adjusted index (orange line) are generally quite close in value throughout the series except for a slight tail off at the end. This would imply that the quality of replacement PCs isn't drastically different to the original product. A possible theory for why this might be the case is that PC sales have been in decline for a few years now as people tend to favour other devices such as laptops and tablets. As a result, it is thought that manufacturers may not be developing PCs to quite the same extent as in the past as the demand for them is dwindling. The resultant IQI values are not too dissimilar to those for food items and suggest that hedonics may be better off being used on other items instead.

Table B4: IQI for smartphones



9. Smartphones are also a hedonics item and the number of quality changes is about standard for an item of its nature.
10. Large differences between the unadjusted and adjusted indices were to be expected for this item as the rate of product churn is high and often the replacement smartphone has a number of different characteristics to the original item (eg. the screensize, memory capacity, camera quality etc. could all be very different as well as the outlet or even brand of the phone). The large differences between the adjusted and unadjusted indices has resulted in rather large IQI values (nearly 140 in some cases) but it is felt that the quality difference between smartphones could be as high as this as they are being developed rapidly meaning replacements are often of higher quality. The high product churn and rapid developments suggest that the use of hedonics on this item is justified.

Table B5: IQI for digital lens cameras



11. The number of quality changes for digital lens cameras may seem low at first glance but considering the sample size is only 16, the proportion of imputation is actually quite high.
12. The standard reference index (blue line) seems to indicate that the price of cameras in the sample generally increases over time, which hints that the replacement cameras are of higher quality to the original items. This is probably what one would expect for a digital item. In contrast, the Quality Adjusted Index remains very flat throughout indicating that, once the impact of quality has been removed, there is not much price change in cameras over the course of the year. While this may be the case, and it is possible that the quality of a camera increases by 27%, one would think that for a digital item like cameras that were would be more of a price change over time but this isn't reflected much in the quality adjusted index.
13. An example that may explain why this was the case is that in June 17, two of the cheapest cameras were replaced by non-comparable replacements that had much higher prices. Because the two cameras had been the cheapest ones, the new base price that was imputed was much greater than the original base price. However, because the actual price of the new cameras was much higher than before, the price relative (current price divided by base price) remained at a similar value to before. In this instance, the Quality Adjusted Index has then taken almost the entire price difference between the original and replacement products to be a quality change and almost none of it to be a price change. CMI may therefore be overestimating quality change for this item. This is especially the case when the sample size of this item is so low as it means that once a lot of replacements have been made, the majority of the sample is based on imputed values.
14. The resultant IQI graph shows quite large differences between the unadjusted and adjusted indices from May 17 until the end of the series which could be the result of the quality adjusted index overestimating quality change and also the result of not having a large enough sample size to measure pure price change.