Office for National Statistics

{75}

Survey Methodology Bulletin

Spring 2016

Contents

Office for National Statistics
Swyddfa Ystadegau Gwladol

The Survey Methodology Bulletin is primarily produced to inform staff in the Office for National Statistics (ONS) and the wider Government Statistical Service (GSS) about ONS survey methodology work. It is produced by ONS, and ONS staff are encouraged to write short articles about methodological projects or issues of general interest. Articles in the bulletin are not professionally refereed, as this would considerably increase the time and effort to produce the bulletin; they are working papers and should be viewed as such.

The bulletin is published twice a year and is available as a download only from the ONS website.

The mission of ONS is to improve understanding of life in the United Kingdom and enable informed decisions through trusted, relevant, and independent statistics and analysis. On 1 April 2008, under the legislative requirements of the 2007 Statistics and Registration Service Act, ONS became the executive office of the UK Statistics Authority. The Authority's objective is to promote and safeguard the production and publication of official statistics that serve the public good and, in doing so, will promote and safeguard (1) the quality of official statistics, (2) good practice in relation to official statistics, and (3) the comprehensiveness of official statistics. The National Statistician is the principal advisor on these matters.

www.ons.gov.uk

Edited by: Joe Winton

methodology@ons.gov.uk

Standard Errors of movement in the Index of Production

Matthew Mayhew[1], Laura Clarke, Charlie Turner and Joseph Winton

Index Numbers Methodology

Summary

The Office for National Statistics (ONS) produces an estimate of index of output of the production industries in the UK, the Index of Production (IoP). Rather than collect this information from all businesses in the production industries, ONS collects data from a sample. This inherently introduces some variation into the estimates produced from this data. This paper describes the method used to produce a Standard Error for the movement of the IoP using the same technique that was used by the ONS for estimation of the variance for the growth of other index outputs such as the Retail Sales Index [1], the House Price Index [2], and Average Weekly Earnings [3]. An indicative standard error for the 12-month growth in the IoP, for the period January to September 2014 is 0.42 percentage points.

1. What is the Index of Production?

The IoP is a monthly indicator measuring the growth in the output of the production industries for the United Kingdom. It is a key economic indicator and is an early measure of economic activity. It is also a component of the output approach to the measuring Gross Domestic Product - GDP(O)[2]. In 2015 the IoP represented 14.9% of GDP(O). It is constructed by taking the total turnover collected by the Monthly Business Survey, MBS, and deflating the total turnover by a weighted combination of the Producer Price Index output series, (PPI) and the Export Price Index (EPI), so that both domestic and foreign sales of products produced in the UK are accounted for. There are some industries where the MBS collects volume data.

For the Manufacturing Sector the data that is obtained through the MBS is the value[3] of the items produced. This is because value data is easier to obtain as the business' accounts would record the turnover from invoices. Ideally for the Index of Production volume data would be collected from the businesses, however volume is hard to collect as information for individual products is too much burden for most businesses to provide, as a compromised the best information that can be readily obtain is the total value of the products. When value is obtained, it needs to be deflated to obtain volume data, as follows: $Q^t = V^t / P^t$; here Q is the volume index, V is value index and P is price index. This is to remove the effect of the change in prices from the change in value.

[1] matthew.mayhew@ons.gov.uk

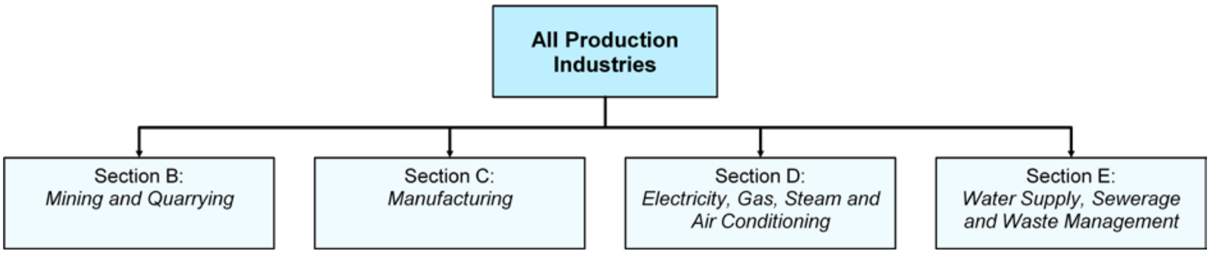
[2] http://www.ons.gov.uk/ons/rel/iop/index-of-production/october-2015/stb-iop-oct-15.html#tab-Gross-domestic- product{GDP{impact-and-components

[3] This article will interchange the use of turnover and value, as they are the same idea in this context

Other SIC sectors, such as mining industries etc, it is easier to obtain volume as they would record the amount they mined in a given period. The different construction of the IoP for volume and value indices leads to different calculations of variance as shown in Section 2.

The Index of Production covers businesses that are classified as part of the production industries according to the Standard Industrial Classification (SIC). The IoP publishes estimates for the growth of production for all production industries and for four sectors of production from SIC, these are detailed in Figure 1.

Figure 1: Structure of the Index of production



What is a Standard Error?

In an ideal world, to calculate the IoP growths, data would be needed from every business in each industry and for each product those businesses produce - approximately 15 million businesses. This would allow us to calculate the exact growth in the production industry for the UK. However, this is not possible as it would be costly and time consuming to collect the data for both ONS and the businesses involved, meaning an increase in the delay in obtaining the data and publishing the IoP growths as well as the burden on the businesses. Instead, a sample of approximately 6,000 businesses is taken, and is used to calculate the IoP. This estimate is dependent on the sample taken, as a different sample would produce a different estimate. The difference between the population value (using all 140,000 in the population.) and the sampling values is called the sampling error, and is unknown, since the population value is unknown, though an estimate of the typical error size, known as the standard error can be estimated from the sample.

The standard error of an index movement is a measure of the spread of possible estimates of that movement likely to be obtained when taking a range of different samples of the same size. The standard error of an estimate, in this case the IoP growths, is a measure of the accuracy of the estimate. The smaller the value of the standard error the more accurate the estimate is and more confidence is given to the estimate of actually representing the population value of IoP growth. The variance of an estimate is equal to the square of the standard error, and measures the spread of the data.

Why a new method has been devised?

The Standard Error of the movement of the Index of Production have previously been produced by the ONS [4], using the same estimation technique outline in section 2. he previous work identified four contributions to the standard error, they are as follows:

1. Monthly Production Inquiry – turnover of production industry
2. Quarterly Stocks Inquiry – the adjustments for the level of stock an industry hold
3. Current Period Deflators – the variation in the Producer Price Index and Export Price Index for the current period
4. Lagged Deflators – these deflate for the stock adjustments.

Following improvements in 2011, and the introduction of the Monthly Business Survey (MBS) in 2010, the IoP only uses inputs of turnovers from the MBS and the deflators, the Producer Price Index and the Export Price Index. Another is the levels and aggregation structure at which standard errors are calculated and reported; standard errors were calculated at SIC03 4-digit level and then aggregated to 2-digit SIC03, Major Industrial Grouping and the all industry level; the current work is calculated at SIC07 2-digit level and then aggregated to SIC07 Sections and the All Product Level.

2. Method of Estimation

The variance of an index number is complex as no exact variance formula exists for the ratio of two random variables, and since the growth in an Index Number is essentially the ratio of two index numbers, the estimation becomes even more complicated. Despite this, there have been methods devised to estimate the variance, and the one chosen for this paper, and other similar outputs released by the ONS, is that of Taylor Linearisation. This method aims to provide a linear approximation to the variance function through the use of a Taylor Series, so using this method our estimator is a linear combination of the variances of the inputs. Therefore the variance of the growth of the IoP is dependent on the variance of its inputs.

Let θ be a vector of parameters and $h(\theta)$ be a function of the parameters, this is also a vector. Then the variance of r^{th} element is:

$$Var(h_r) = \sum_i \left(\frac{\partial h_r}{\partial \hat{\theta}_i}\right)^2 Var(\hat{\theta}_i) + \sum_i \sum_{j \neq i} \left(\frac{\partial h_r}{\partial \hat{\theta}_i}\right)\left(\frac{\partial h_r}{\partial \hat{\theta}_j}\right) Cov(\hat{\theta}_i, \hat{\theta}_j) \qquad [1]$$

Where $\hat{\theta}_i$ is an estimator of θ_i. There are two special cases of this, where h is a product of two parameters, and where h is a quotient of two parameters.

$$Var(AB) = Var(A)Var(B) + Var(A)\mathrm{E}(B)^2 + \mathrm{E}(A)^2 Var(B) \qquad [2]$$

$$Var\left(\frac{A}{B}\right) = \left(\frac{A}{B}\right)^2 \left(\frac{Var(A)}{A^2} + \frac{Var(B)}{B^2} - \frac{2Cov(A,B)}{AB}\right) \qquad [3]$$

The formula for the product is only valid if the two random variables are uncorrelated.

The growth in the Index of Production for a volume industry is:

$$\frac{\hat{I}_{ht}}{\hat{I}_{hs}} = \frac{\hat{V}_{ht}}{\hat{V}_{hs}} \qquad [4]$$

Where V_{ht} is the volume of production in period t, for the volume industry h. Using the variance of a ratio formula, the following estimate of the variance of growth in an IoP volume industry is obtained:

$$Var\left(\frac{\hat{V}_{ht}}{\hat{V}_{hs}}\right) = \left(\frac{\hat{V}_{ht}}{\hat{V}_{hs}}\right)^2 \left(\frac{Var(\hat{V}_{ht})}{\hat{V}_{ht}^2} + \frac{Var(\hat{V}_{hs})}{\hat{V}_{hs}^2} - 2\frac{Cov(\hat{V}_{hs},\hat{V}_{ht})}{\hat{V}_{hs}\hat{V}_{ht}}\right) \qquad [5]$$

The formula for a turnover industry is more complicated, but a Taylor linearisation can be performed. The growth from month s to month t is:

$$\frac{\hat{I}_{ht}}{\hat{I}_{hs}} = \frac{\hat{C}_{ht}\sum_{p\in I}\frac{1}{\hat{D}_{pt}}\frac{\hat{w}_{hpt}}{\sum_{q\in I}\hat{w}_{hqt}}}{\hat{C}_{hs}\sum_{p\in I}\frac{1}{\hat{D}_{ps}}\frac{\hat{w}_{hps}}{\sum_{q\in I}\hat{w}_{hqs}}} \qquad [6]$$

Where C_{ht} is the turnover in period t, for the turnover industry h; D_{pt} is the deflator for product p in period t and w_{hpt} is the weight of product p in industry h in period t. The estimate of the variance of growth in an IoP turnover industry from period s to period t is as follows[4]:

$$\begin{aligned} Var\left(\frac{\hat{I}_{ht}}{\hat{I}_{hs}}\right) = &\left(\frac{\partial f}{\partial \hat{C}_{Eht}}\right)^2 Var(\hat{C}_{Eht}) + \left(\frac{\partial f}{\partial \hat{C}_{Ehs}}\right)^2 Var(\hat{C}_{Ehs}) \\ &+ \left(\frac{\partial f}{\partial \hat{C}_{Hht}}\right)^2 Var(\hat{C}_{Hht}) + \left(\frac{\partial f}{\partial \hat{C}_{Hht}}\right)^2 Var(\hat{C}_{Hht}) \\ &+ \frac{\partial f}{\partial \hat{C}_{Eht}}\frac{\partial f}{\partial \hat{C}_{Ehs}} Cov(\hat{C}_{Eht},\hat{C}_{Ehs}) \\ &+ \frac{\partial f}{\partial \hat{C}_{Hht}}\frac{\partial f}{\partial \hat{C}_{Hhs}} Cov(\hat{C}_{Hht},\hat{C}_{Hhs}) \\ &+ \left(\frac{\partial f}{\partial \hat{D}_{ht}}\right)^2 Var(\hat{D}_{ht}) + \left(\frac{\partial f}{\partial \hat{D}_{hs}}\right)^2 Var(\hat{D}_{hs}) \end{aligned} \qquad [7]$$

Where f is the growth of IoP as defined by equation 6.

[4] Detailed Derivation in Annex A

3. Results

Standard Errors of the movement of IoP have been calculated at two frequencies of growth: month on previous month (monthly growths); and month on same month a year ago (12-month growths). Standard Errors of the movement have been calculated at the overall IoP, or All Industry level as well as at the section level.

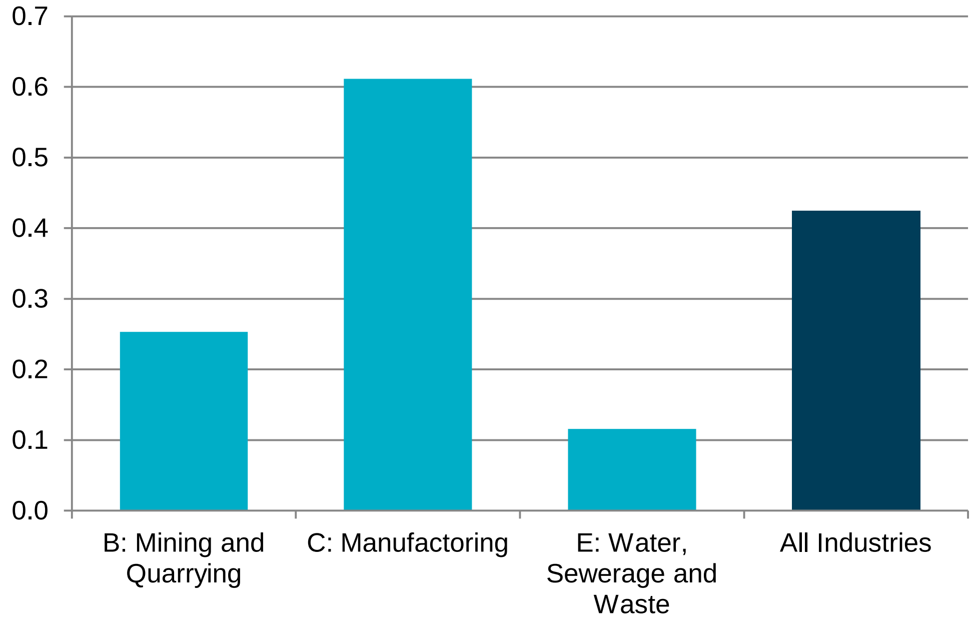
NOTE: Calculations have not been made on Section D as the data obtained for that section is a census of the industries, hence standard errors are zero.

3.1. 12-Month Growths

The twelve month growths show the change in the amount of production output between the current month and the same month in the previous year, and should not be affected by seasonal effects. Figure 2 presents the median standard error in the 12 month growth of IoP by sector. The standard error in the All Industries growth is driven predominantly by the large standard error in the Manufacturing sector which accounts for approximately 70% of the Index of Production.

The standard error is useful for interpreting the accuracy of estimate, where if a 95% confidence interval around the estimate contains the value 0 then the estimate is not significantly different from 0. Table 1 shows the growth and estimate of the standard errors for this period and Figure 3 presents the growths in charts with 95% confidence intervals.

Figure 2: Median Standard Error in the 12 month growth of the Index of Production, January – September 2014

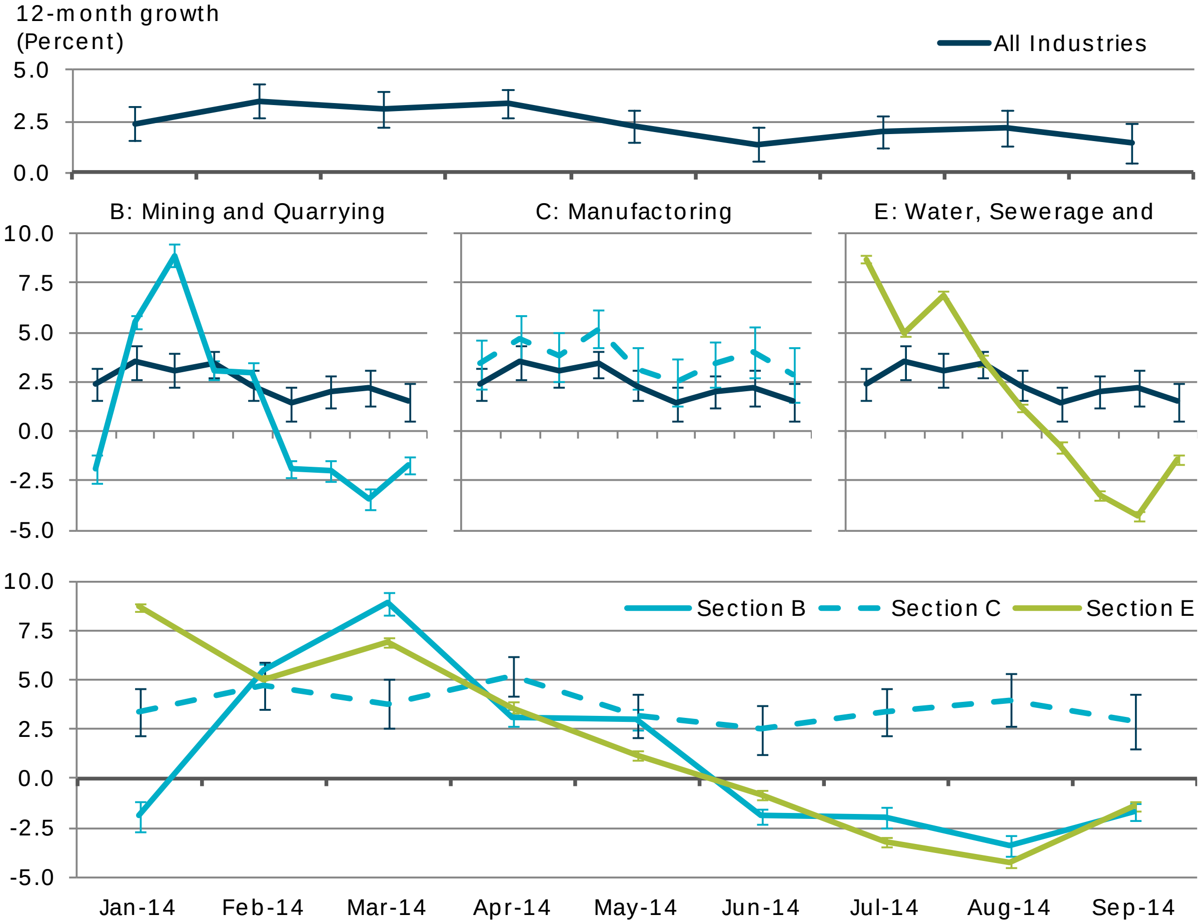


As it can be seen all the growths are significant for all levels of aggregation, therefore there is evidence that there was growth or decline in the production industries at these periods of time. Section C has a higher standard error than the other sections; as this section is comprises of turnover industries, Manufacturing is affected by variation in the price indices used as deflators; if there is a larger variation in the prices of the products produced by Manufacturing sector then the variation in the growth of output for this section will be larger as well.

Table 1: Standard Errors of the 12 month growth in the Index of Production

Period	All Industries 12 month growth	All Industries SE	Section B 12 month growth	Section B SE	Section C 12 month growth	Section C SE	Section E 12 month growth	Section E SE
Jan-14	2.40	0.43	-1.90	0.38	3.40	0.62	8.70	0.10
Feb-14	3.50	0.43	5.50	0.17	4.70	0.61	5.00	0.11
Mar-14	3.10	0.44	8.90	0.29	3.80	0.64	6.90	0.11
Apr-14	3.40	0.34	3.10	0.24	5.20	0.49	3.60	0.13
May-14	2.30	0.39	3.00	0.25	3.20	0.56	1.20	0.11
Jun-14	1.40	0.43	-1.90	0.21	2.50	0.62	-0.80	0.13
Jul-14	2.00	0.41	-2.00	0.27	3.40	0.60	-3.20	0.12
Aug-14	2.20	0.45	-3.40	0.26	4.00	0.66	-4.30	0.12
Sep-14	1.50	0.49	-1.70	0.22	2.90	0.71	-1.40	0.11

Figure 3 Twelve month growth in the Index of Production with 95% Confidence Interval, by Industry



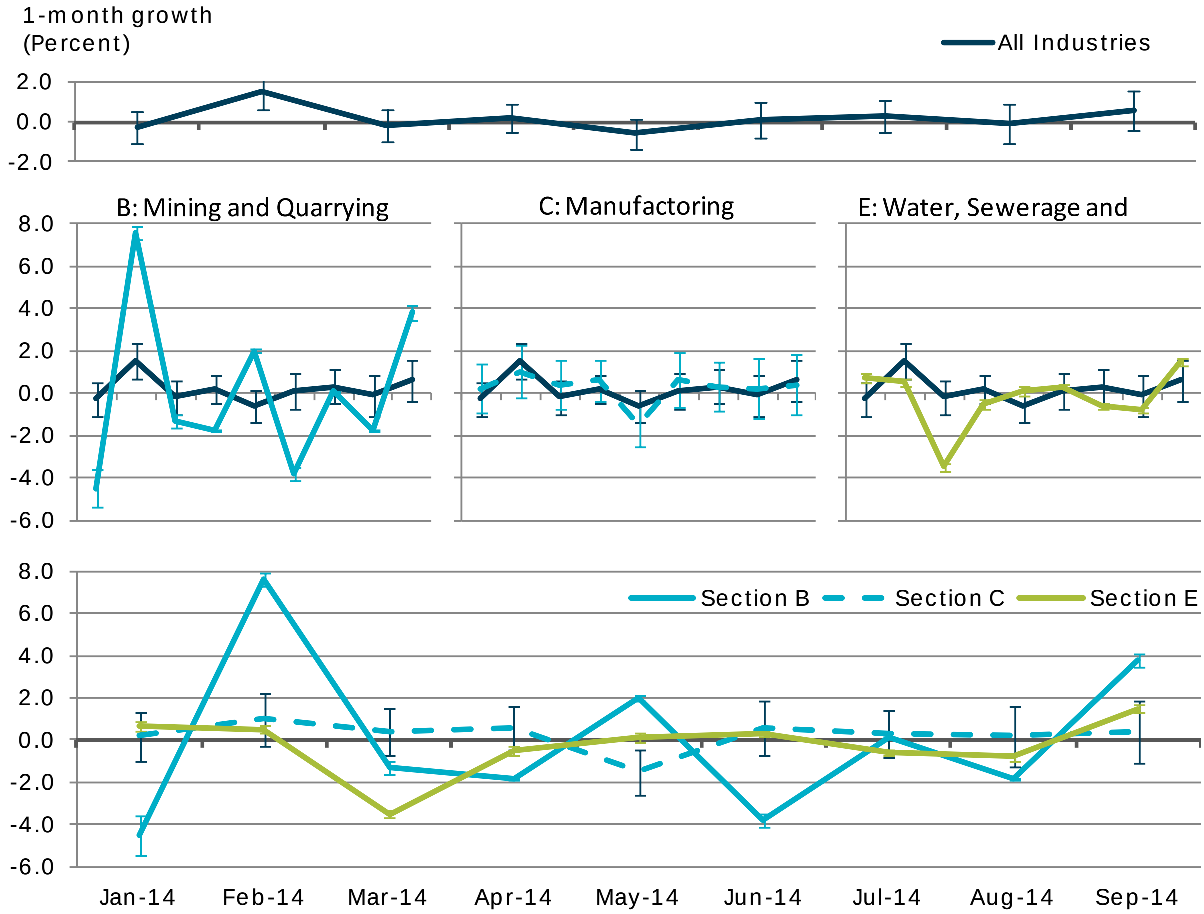
3.2. Monthly Growths

The monthly growths show the change in the amount of production output between the current month and the previous month, these are more often than not smaller than the 12 month growths, and are often closer to zero. Table 2 shows the monthly growths and estimate of its standard errors for September 2014.

Table 2: Standard Errors of the monthly growth in the Index of Production, January - September 2014

	All Industries		Section B		Section C		Section E	
Period	12 month growth	SE	12 month growth	SE	12 month growth	SE	12 month growth	SE
Jan-14	-0.30	0.41	-4.50	0.46	0.20	0.59	0.70	0.12
Feb-14	1.50	0.43	7.60	0.15	1.00	0.63	0.50	0.09
Mar-14	-0.20	0.40	-1.30	0.15	0.40	0.58	-3.50	0.10
Apr-14	0.20	0.36	-1.80	0.02	0.60	0.51	-0.50	0.11
May-14	-0.60	0.37	2.00	0.05	-1.50	0.54	0.10	0.11
Jun-14	0.10	0.45	-3.80	0.16	0.60	0.65	0.30	0.07
Jul-14	0.30	0.40	0.10	0.02	0.30	0.58	-0.60	0.07
Aug-14	-0.10	0.50	-1.80	0.03	0.20	0.73	-0.80	0.08
Sep-14	0.60	0.51	3.80	0.17	0.40	0.74	1.50	0.09

Figure 4 One month growth in the Index of Production with 95% Confidence Interval, by Industry



For the all industry level all of the growths are not significantly different from zero accept of February, this is mainly due to the growths themselves being closer to zero. The growth for Section B is significant as the growths are higher than for the other sections, as well as having smaller standard errors. Though in July, the growth is close to zero but the standard error is small enough that the confidence interval does not include zero. Section E growths are all significant accept for May where the groeth is closer to zero than the other months. Section E also has thinner confidence intervals the other sections in each most likely due to small standard errors whose values are close to each other. The growths with confidence intervals for January to September are shown in Figure 4.

An interesting observation is that the confidence intervals for Section C always overlap with the All Industries confidence interval, this may be caused by the fact that Section C has the highest weight when aggregating the IoP. The other sections don't often overlap with the All Industries; this might also be due to the weights given to those sections.

References

[1] Joseph Winton and Jeff Ralph. Measuring the accuracy of the retail sales index. 2011.(http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/retail-sales/measuring-the-accuracy-of-the-retail-sales-index.pdf [Accessed 25/05/16])

[2] Robert O'Neill, Gareth Clews and Jeff Ralph. The Methodology Used to Estimate the Standard Errors of Movement in the UK House Price Index. 2015. (http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-73/survey-methodology-bulletin-73---spring-2015.pdf [Accessed 25/05/16])

[3] Gareth Clews, Ria Sanderson and Jeff Ralph. The Method Used to Estimate the Standard Errors of Movement for UK Average Weekly Earnings. 2014. (http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-72/survey-methodology-bulletin--no-72--spring-2014.pdf [Accessed 25/05/16)

[4] John Wood, Markus G. Šova, Neil Parkin and Robert D. Bucknall. Estimation of Standard Errors for the UK Index of Production. (http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-59/survey-methodology-bulletin-59---sept-2006.pdf [accessed 25/05/16])

Annex A

Estimation of the variance for the Index of Production

For a turnover industry, the growth from month s to month t is

$$
\begin{aligned}
\frac{\hat{I}_{ht}}{\hat{I}_{hs}} &= \frac{\sum_{p\in\mathcal{I}} \frac{\hat{C}_{ht}}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\sum_{p\in\mathcal{I}} \frac{\hat{C}_{hs}}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} \\
&= \frac{\hat{C}_{ht} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\hat{C}_{hs} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}}
\end{aligned}
$$

This formula obscures the fact that if p corresponds to a turnover industry, the product deflators $\hat{D}_{ps}$ and $\hat{D}_{pt}$ depend on weights derived from the export and home turnovers $\hat{C}'_{Ept}$ and $\hat{C}'_{Hpt}$ for the industry corresponding to p. Recall also, that the turnover for industry h is the sum of the export and import turnovers, i.e. $\hat{C}_{ht} = \hat{C}_{Eht} + \hat{C}_{Hht}$. Let's consider the RHS of the above equation as a function

$$
f\left(\hat{C}_{Eht}, \hat{C}_{Ehs}, \hat{C}_{Hht}, \hat{C}_{Hhs}, (\hat{D}_{pt})_{p\in\mathcal{I}}, (\hat{D}_{ps})_{p\in\mathcal{I}}, \left(\frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}\right)_{p\in\mathcal{I}}, \left(\frac{\hat{w}_{hqs}}{\sum_{q\in\mathcal{I}} \hat{w}_{hps}}\right)_{p\in\mathcal{I}}\right).
$$

Remember that the deflators are function of the turnovers. To use Taylor linearisation, the derivatives need to be calculated. Taking the derivative with respect to the deflators, the following is obtained

$$
\frac{\partial f}{\partial \hat{D}_{pt}} = -\frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\left(\frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hpt}}\right) \frac{1}{\hat{D}^2_{pt}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)}
$$

which holds for all $p \in \mathcal{I}$.

$$
\frac{\partial f}{\partial \left(\frac{\hat{w}_{hpt}}{\sum_{q\mathcal{I}} \hat{w}_{hqt}}\right)} = \frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\frac{1}{\hat{D}_{pt}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)}
$$

$$
\frac{\partial f}{\partial \hat{C}_{Eht}} = \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\hat{C}_{hs} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} + \frac{\partial f}{\partial \hat{D}'_{ht}} \frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Eht}}
$$

$$
\frac{\partial f}{\partial \hat{C}_{Hht}} = \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\hat{C}_{hs} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} + \frac{\partial f}{\partial \hat{D}'_{ht}} \frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Hht}}
$$

The last two expressions require the derivatives $\frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Eht}}$ and $\frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Hht}}$ respectively. The deflator is given by

$$\hat{D}_{pt} = \frac{1}{\hat{d}_{p0}} \frac{1}{\left(\frac{\hat{C}'_{Ept}}{(\hat{C}'_{Ept}+\hat{C}'_{Hpt})\widehat{EPI}_{pt}} + \frac{C'_{Hpt}}{(\hat{C}'_{Ept}+\hat{C}'_{Hpt})\widehat{PPI}_{pt}} \right)}$$

which can be rearranged as

$$\hat{D}_{pt} = \frac{1}{\hat{d}_{p0}} \frac{(\hat{C}'_{Ept} + \hat{C}'_{Hpt})\widehat{EPI}_{pt}\widehat{PPI}_{pt}}{(\hat{C}'_{Ept}\widehat{PPI}_{pt} + \hat{C}'_{Hpt}\widehat{EPI}_{pt})}$$

Swapping again between products and industries to focus on industry h, the derivative is:

$$\frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Eht}} = \frac{1}{\hat{d}_{h0}} \frac{\hat{C}'_{Hht}\widehat{EPI}_{ht}\widehat{PPI}_{ht}(\widehat{EPI}_{ht} - \widehat{PPI}_{ht})}{(\hat{C}'_{Ept}\widehat{PPI}_{pt} + \hat{C}'_{Hpt}\widehat{EPI}_{pt})^2}$$

$$\frac{\partial \hat{D}'_{ht}}{\partial \hat{C}_{Hht}} = \frac{1}{\hat{d}_{h0}} \frac{\hat{C}'_{Eht}\widehat{EPI}_{ht}\widehat{PPI}_{ht}(\widehat{EPI}_{ht} - \widehat{PPI}_{ht})}{(\hat{C}'_{Ept}\widehat{PPI}_{pt} + \hat{C}'_{Hpt}\widehat{EPI}_{pt})^2}$$

So from this the derivates with respect to export turnover is

$$\frac{\partial f}{\partial \hat{C}_{Eht}} = \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\hat{C}_{hs} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} - \frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\left(\frac{\hat{w}_{hht}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}\right) \frac{1}{\hat{D}^2_{ht}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)} \frac{1}{\hat{d}_{h0}} \frac{\hat{C}'_{Hht}\widehat{EPI}_{ht}\widehat{PPI}_{ht}(\widehat{EPI}_{ht} - \widehat{PPI}_{ht})}{(\hat{C}'_{Ept}\widehat{PPI}_{pt} + \hat{C}'_{Hpt}\widehat{EPI}_{pt})^2}$$

and with respect to home turnover

$$\frac{\partial f}{\partial \hat{C}_{Hht}} = \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\hat{C}_{hs} \sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} - \frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\left(\frac{\hat{w}_{hht}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}\right) \frac{1}{\hat{D}^2_{ht}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)} \frac{1}{\hat{d}_{h0}} \frac{\hat{C}'_{Eht}\widehat{EPI}_{ht}\widehat{PPI}_{ht}(\widehat{EPI}_{ht} - \widehat{PPI}_{ht})}{(\hat{C}'_{Ept}\widehat{PPI}_{pt} + \hat{C}'_{Hpt}\widehat{EPI}_{pt})^2}.$$

Now for all the derivatives with respect to month s. As all variables with respect to s are in the denominator the expressions are going to be uglier.

$$\frac{\partial f}{\partial \hat{C}_{Ehs}} = -\frac{\hat{C}_{ht}}{\hat{C}^2_{hs}} \left(\frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} \right) + \frac{\partial f}{\partial \hat{D}'_{hs}} \frac{\partial \hat{D}'_{hs}}{\partial \hat{C}_{Ehs}}$$

$$\frac{\partial f}{\partial \hat{C}_{Hhs}} = -\frac{\hat{C}_{ht}}{\hat{C}^2_{hs}} \left(\frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} \right) + \frac{\partial f}{\partial \hat{D}'_{hs}} \frac{\partial \hat{D}'_{hs}}{\partial \hat{C}_{Hhs}}$$

$$\frac{\partial f}{\partial \hat{D}_{ps}} = -\frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}} \frac{1}{\hat{D}^2_{ps}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)^2}$$

$$\frac{\partial f}{\partial \left(\frac{\hat{w}_{hps}}{\sum_{p\in\mathcal{I}} \hat{w}_{hqs}}\right)} = -\frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{p\in\mathcal{I}} \hat{w}_{hqt}} \frac{1}{\hat{D}_{ps}}}{\left(\frac{\hat{w}_{hps}}{\sum_{p\in\mathcal{I}} \hat{w}_{hqs}}\right)^2}$$

Hence the following is obtained:

$$\frac{\partial f}{\partial \hat{C}_{Ehs}} = -\frac{\hat{C}_{ht}}{\hat{C}^2_{hs}} \left(\frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}}}{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}} \right) - \frac{\hat{C}_{ht}}{\hat{C}_{hs}} \frac{\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{pt}} \frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqt}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}} \frac{1}{\hat{D}^2_{ps}}}{\left(\sum_{p\in\mathcal{I}} \frac{1}{\hat{D}_{ps}} \frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}} \hat{w}_{hqs}}\right)^2}$$

$$\times \frac{1}{\hat{D}_{h0}} \frac{\hat{C}'_{Hhs}\widehat{EPI}_{hs}\widehat{PPI}_{hs}(\widehat{EPI}_{hs}-\widehat{PPI}_{hs})}{(\hat{C}'_{Eps}\widehat{PPI}_{ps}+\hat{C}'_{Hps}\widehat{EPI}_{ps})^2}$$

$$\frac{\partial f}{\partial \hat{C}_{Hhs}} = -\frac{\hat{C}_{ht}}{\hat{C}_{hs}^2}\left(\frac{\sum_{p\in\mathcal{I}}\frac{1}{\hat{D}_{pt}}\frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}}\hat{w}_{hqt}}}{\sum_{p\in\mathcal{I}}\frac{1}{\hat{D}_{ps}}\frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}}\hat{w}_{hqs}}}\right) - \frac{\hat{C}_{ht}}{\hat{C}_{hs}}\frac{\sum_{p\in\mathcal{I}}\frac{1}{\hat{D}_{pt}}\frac{\hat{w}_{hpt}}{\sum_{q\in\mathcal{I}}\hat{w}_{hqt}}\frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}}\hat{w}_{hqs}}\frac{1}{\hat{D}_{ps}^2}}{\left(\sum_{p\in\mathcal{I}}\frac{1}{\hat{D}_{ps}}\frac{\hat{w}_{hps}}{\sum_{q\in\mathcal{I}}\hat{w}_{hqs}}\right)^2}$$

$$\times \frac{1}{\hat{D}_{h0}} \frac{\hat{C}'_{Ehs}\widehat{EPI}_{hs}\widehat{PPI}_{hs}(\widehat{EPI}_{hs}-\widehat{PPI}_{hs})}{(\hat{C}'_{Eps}\widehat{PPI}_{ps}+\hat{C}'_{Hps}\widehat{EPI}_{ps})^2}$$

These can then be placed in the Taylor Linearisation formula.

Variance of the Deflator

Recall that the deflator is defined as

$$\hat{D}_{pt} = \frac{\hat{D}_{pt}}{\hat{D}_{p0}}.$$

Here there is a ratio of two variables $\hat{D}_{pt}$ and $\hat{D}_{p0}$ after some manipulation and simplification the numerator can be written as

$$\frac{C'_{Ep0}}{\hat{C}'_{Tp0}\widehat{EPI}_{p0}} + \frac{C'_{Hp0}}{\hat{C}'_{Tp0}\widehat{PPI}_{p0}}$$

where the subscript T represents the sum of export and home turnovers, the denominator is the same with the time period changed to t. For the variance calculation, the notation is simplified as follows $A = \frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}} + \frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}$.

$$\begin{aligned}
\text{Var}(A) &= \text{Var}\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}} + \frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)\\
&= \text{Var}\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}\right) + \text{Var}\left(\frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right) + 2\text{Cov}\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}, \frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)\\
&= \left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{E0})}{\hat{C}_{E0}^2} + \frac{\text{Var}(\hat{C}_{T0}\widehat{EPI}_{p0})}{(\hat{C}_{T0}\widehat{EPI}_{p0})^2} - 2\text{Cov}(\hat{C}_{E0}, \hat{C}_{T0}\widehat{EPI}_{p0})\right)\\
&+ \left(\frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{H0})}{\hat{C}_{H0}^2} + \frac{\text{Var}(\hat{C}_{T0}\widehat{PPI}_{p0})}{(\hat{C}_{T0}\widehat{PPI}_{p0})^2} - 2\text{Cov}(\hat{C}_{H0}, \hat{C}_{T0}\widehat{PPI}_{p0})\right)\\
&+ 2\text{Cov}\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}, \frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)\\
&= \left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{E0})}{\hat{C}_{E0}^2} + \frac{\text{Var}(\hat{C}_{T0})\text{Var}(\widehat{EPI}_{p0}) + \text{Var}(\hat{C}_{T0})\mathbb{E}(\widehat{EPI}_{p0})^2 + \mathbb{E}(\hat{C}_{T0})^2\text{Var}(\widehat{EPI}_{p0})}{(\hat{C}_{T0}\widehat{EPI}_{p0})^2} - 2\text{Cov}(\hat{C}_{E0}, \hat{C}_{T0}\widehat{EPI}_{p0})\right)\\
&+ \left(\frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{H0})}{\hat{C}_{H0}^2} + \frac{\text{Var}(\hat{C}_{T0})\text{Var}(\widehat{PPI}_{p0}) + \text{Var}(\hat{C}_{T0})\mathbb{E}(\widehat{PPI}_{p0})^2 + \mathbb{E}(\hat{C}_{T0})^2\text{Var}(\widehat{PPI}_{p0})}{(\hat{C}_{T0}\widehat{PPI}_{p0})^2} - 2\text{Cov}(\hat{C}_{H0}, \hat{C}_{T0}\widehat{PPI}_{p0})\right)\\
&+ 2\text{Cov}\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}, \frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)
\end{aligned}$$

There is reasonable assumption that *"Estimators from separate surveys are independent"* so the covariance terms are zero and the variance of A is

$$\begin{aligned}
\text{Var}(A) = &\left(\frac{\hat{C}_{E0}}{\hat{C}_{T0}\widehat{EPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{E0})}{\hat{C}_{E0}^2} + \frac{\text{Var}(\hat{C}_{T0})\text{Var}(\widehat{EPI}_{p0}) + \text{Var}(\hat{C}_{T0})\mathbb{E}(\widehat{EPI}_{p0})^2 + \mathbb{E}(\hat{C}_{T0})^2\text{Var}(\widehat{EPI}_{p0})}{(\hat{C}_{T0}\widehat{EPI}_{p0})^2}\right)\\
&+ \left(\frac{\hat{C}_{H0}}{\hat{C}_{T0}\widehat{PPI}_{p0}}\right)^2\left(\frac{\text{Var}(\hat{C}_{H0})}{\hat{C}_{H0}^2} + \frac{\text{Var}(\hat{C}_{T0})\text{Var}(\widehat{PPI}_{p0}) + \text{Var}(\hat{C}_{T0})\mathbb{E}(\widehat{PPI}_{p0})^2 + \mathbb{E}(\hat{C}_{T0})^2\text{Var}(\widehat{PPI}_{p0})}{(\hat{C}_{T0}\widehat{PPI}_{p0})^2}\right)
\end{aligned}$$

The variance of the denominator is the same as for the numerator but replacing the 0 subscript with t or s depending on whether the variance is needed for the deflator in period t or in period s.

The final equation for the growth of the index of production from period s to period t is as follows.

$$
\begin{aligned}
\mathsf{Var}\left(\frac{\hat{I}_{ht}}{\hat{I}_{hs}}\right) = {} & \left(\frac{\partial f}{\partial \hat{C}_{Eht}}\right)^2 \mathsf{Var}(\hat{C}_{Eht}) + \left(\frac{\partial f}{\partial \hat{C}_{Ehs}}\right)^2 \mathsf{Var}(\hat{C}_{Ehs}) \\
& + \left(\frac{\partial f}{\partial \hat{C}_{Hht}}\right)^2 \mathsf{Var}(\hat{C}_{Hht}) + \left(\frac{\partial f}{\partial \hat{C}_{Hhs}}\right)^2 \mathsf{Var}(\hat{C}_{Hhs}) \\
& + \frac{\partial f}{\partial \hat{C}_{Eht}} \frac{\partial f}{\partial \hat{C}_{Ehs}} \mathsf{Cov}(\hat{C}_{Eht}, \hat{C}_{Ehs}) \\
& + \frac{\partial f}{\partial \hat{C}_{Hht}} \frac{\partial f}{\partial \hat{C}_{Hhs}} \mathsf{Cov}(\hat{C}_{Hht}, \hat{C}_{Hhs}) \\
& + \left(\frac{\partial f}{\partial \hat{D}_{ht}}\right)^2 \mathsf{Var}(\hat{D}_{ht}) + \left(\frac{\partial f}{\partial \hat{D}_{hs}}\right)^2 \mathsf{Var}(\hat{D}_{hs})
\end{aligned}
$$

Note that the effect of the weights is assumed to be negligible so has been ignored in the final equation.

Investigating attrition on the Labour Force Survey

Andrea Lacey and Matt Greenaway
Survey Methodology and Statistical Computing Division, ONS

Abstract

With response rates for social surveys continuing to fall, the issue of responders 'dropping out' of longitudinal or rotating panel surveys – known as 'attrition' - is becoming increasingly important. A recommendation of the National Statistics Quality Review (NSQR) of the Labour Force Survey (LFS) was that further research be carried out to assess more fully the process of attrition and its consequences for headline LFS estimates. The focus of this research was to answer three questions:

- What are the respondent characteristics that influence an individual's likelihood to drop out of the LFS?
- Does attrition have an impact on headline LFS estimates?
- What methods may mitigate any attrition bias?

Logistic regression was carried out using the six most influential characteristics which appear to influence an individual's propensity to drop out - age, region, tenure, disability status, ethnicity and household type. Parameter estimates from this model were used to apply a 'sample based' adjustment to the LFS weighting. The impact this had on headline labour market statistics suggests that the attrition described by the model does have a notable but fairly consistent impact on headline LFS point estimates, although investigation over a longer time-period is required to evaluate whether this impact remains consistent under different economic conditions and changing overall non-response patterns.

Further investigations have identified two options which may reduce this attrition bias; utilising a 'sample-based' tenure adjustment in the weighting method, and adjusting the 'rolling data forward' imputation method. The authors are grateful to Debbie Cooper and Fola Ariyibi for their advice and comments throughout this investigation.

1. Introduction

The Labour Force Survey (LFS) is a quarterly survey using a 'rotating panel' design – households, once they enter the sample, continue to be sampled for five consecutive quarters. This design is visualised in figure 1.

In addition to non-response at first contact, rotating panel designs suffer from 'attrition', which we define as individuals not responding at waves 2-5, given that they have responded in a previous wave. Labour Force Survey attrition has been increasing over time, and a recommendation of the LFS National Statistics Quality review (NSQR) (ONS, 2014) was that an investigation be carried out into potential bias caused by attrition and ways of correcting this in LFS estimation.

Figure 1 The structure of the LFS. The coloured horizontal groups represent a 'panel' of individuals, and the vertical groups represent a single quarterly LFS dataset

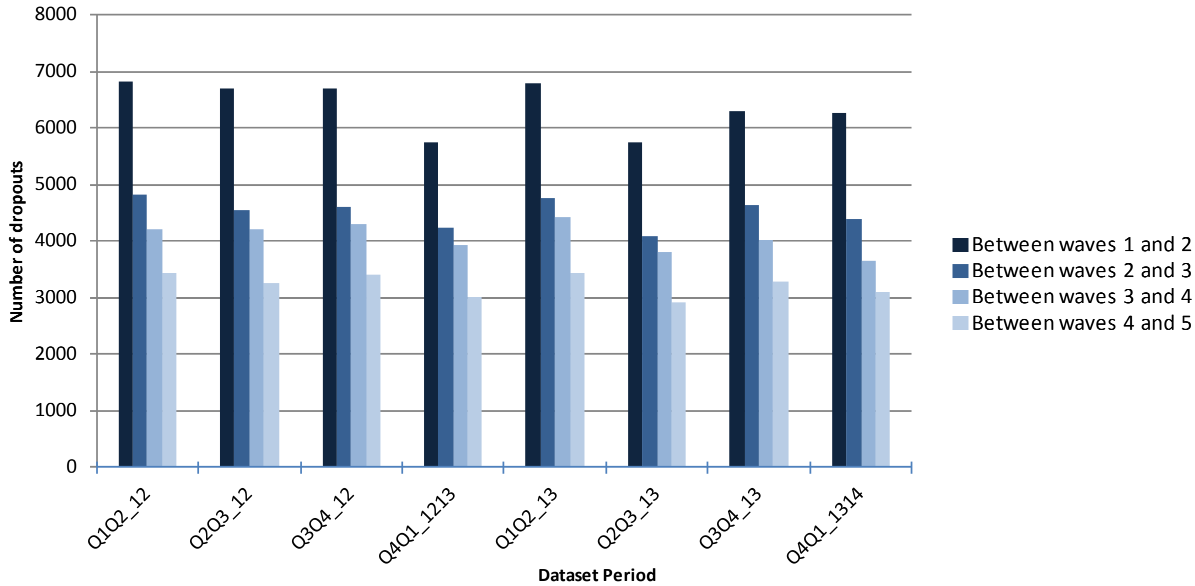
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
				W1	W2	W3	W4
			W1	W2	W3	W4	W5
		W1	W2	W3	W4	W5	
	W1	W2	W3	W4	W5		
W1	W2	W3	W4	W5			
W2	W3	W4	W5				
W3	W4	W5					
W4	W5						
W5							

Attrition is not a simple linear process of individuals responding at wave 1 and then gradually dropping out – individuals can drop out and then respond again at later waves. They can also enter the sample for the first time at waves 2-5; if new individuals move in to a sampled household, these individuals will be captured by the sample (and any individuals who move out will be dropped). We have limited this study to considering two consecutive LFS quarters and evaluating the individuals who drop out between two consecutive waves. We do not consider individuals who re-enter the sample at later waves.

2. Descriptive analysis

Figure 2, below, shows the number of individuals who drop out of the survey between any two consecutive waves, between Q1 2012 and Q1 2014. The greatest number of dropouts is consistently between waves 1 and 2, with progressively fewer dropouts at later waves.

Figure 2 Number of dropouts between quarters, and the waves between which the dropout occurs



A brief review of the literature suggests that around 30 characteristics have been shown to be predictive of attrition and non-response. The census non-response link study (Ashworth et al, 2013) is the most high-profile study; this identified a number of variables associated with wave 1 non-response on the Labour Force Survey. Clarke and Tate (1999) and Kanabar (2013) focus specifically on LFS attrition rather than non-response, but primarily in the context of the 'longitudinal' or 'flows' estimates.

A selection of variables identified by this review are shown below, alongside the attrition rate between two given quarters for different groups.

Table 1: Attrition rates for key variables (average = 25%) between Q2 and Q3, 2013.

Variable Description	Highest attrition rate	Attrition rate	Lowest dropout rate	Dropout rate
Age Band	20-24 year olds	43%	Over 65s	15%
Tenure	Rented accommodation	33%	Owned outright	16%
Region	Inner/Outer London	33%	South West	21%
Household type	Two or more people	46%	One person	21%
Labour Market Status	Unemployed	31%	Inactive	19%
Ethnicity	Mixed/Multiple ethnic groups	46%	White	21%
Marital Status	Single	30%	Widowed	13%
Number of family units	As the number of family units in a household increases, the attrition rate increases			
Sex	No difference (both 25%)			
Time at address	As length of time at address increases, the attrition rate decreases			

The impact of Labour Market Status is particularly notable; 31% of unemployed individuals drop out of the survey between any two consecutive quarters while only 19% of inactive individuals do. While this may to an extent be corrected by the survey weighting as it currently stands (the figures in Table 1 are unweighted), this is a clear indication that attrition propensities may be correlated with key survey outcome variables. However, it is important to emphasise that this table simply reports dropout rates for a number of variables independently – so difference in dropout rates between unemployed and inactive individuals does not control for age, tenure, or any of the other variables in the table. Evaluating the impact of one variable whilst controlling for others requires the use of statistical modeling.

3. Attrition model

The variables identified as important by the exploratory analysis (described above) as having a consistent and significant impact on attrition were used to model attrition in a number of periods. –These variables were - household type, region, age, tenure, ethnicity and disability status. These six variables were used in the final attrition model. Table 2, below, summarises the Wald statistics and p-values for these variables obtained from one period of data using our attrition model.

Table 2: Wald statistics and p-values for final attrition model

Variable	Wald Chi-Square	P value
Region	124.292	<.0001
Tenure	294.765	<.0001
5 Year Age Bands	928.402	<.0001
Household Type	216.307	<.0001
Ethnicity	77.340	<.0001
Disability status	41.180	<.0001

Although we noted in table 1 that the dropout rate varies by labour market status, the labour market status variable is not significant in the model – if included it has a p-value of 0.877. This is consistent with earlier research on LFS attrition, in particular Clarke and Tate (1999), and implies that the likelihood of an individual dropping out of the survey may not be directly influenced by their labour market status after region, tenure, age, household type, ethnicity and disability are controlled for.

It is worth noting that while we can evaluate the impact of labour market status, we cannot account for the influence of changes in labour market status, since we do not have information on time-2 labour market status for drop-outs. We cannot evaluate, for example, whether unemployed individuals who become inactive between two periods are any more or less likely to drop out of the survey between these two periods.

The pseudo R-squared for this model is consistently low, at around 8%. Although the model has identified a number of variables which have a significant impact on attrition, overall it still explaining relatively little of the variation in attrition. The results, presented in section 4, should therefore be treated as describing the impact of the relatively small amount of attrition which we can explain, rather than the entirety of the impact of attrition on LFS estimates.

The odds ratios for all categories of each variable in the model are shown in Appendix 1. Since the model was predicting propensity to stay in the survey, the odds ratios should be interpreted as the multiplicative impact of being in each group on the odds of remaining in the survey. Some key effects are that, holding other variables constant

- Those who own their own home outright are considerably less likely to drop out of the survey, while those who rent are more likely to drop out
- Married couple households are less likely to drop out of the survey, while individuals in households containing multiple family units are much more likely to drop out
- Younger individuals are more likely to drop out of the survey
- White individuals are less likely to drop out of the survey
- The odds of dropping out of the survey vary substantially by region

Many of these findings are consistent with most existing research non-response. However some parameter estimates are surprising; in particular, individuals in single person households or lone parents appear to be more likely to stay in the survey holding other variables constant, a different result from much non-response research.

This suggests that panel survey attrition and initial non-response may be in some respects quite distinct processes.

4. Sample-based weighting based on attrition model

The LFS at present uses a 'population-based' weighting method – each case is assigned a design weight based on the inverse of their probability of selection, and these weights are calibrated to known population totals. We apply a 'sample-based' adjustment to the design weights using attrition probabilities given by the model as described above, and calibrate these adjusted design weights in the same way as in current LFS estimation. Applying this adjustment ensures that those with a lower probability of staying in the survey get a larger weight, reducing attrition bias, although there will be a consequential increase in standard errors.

Figure 3a: Total employment (aged 16+), when an attrition adjustment is applied.

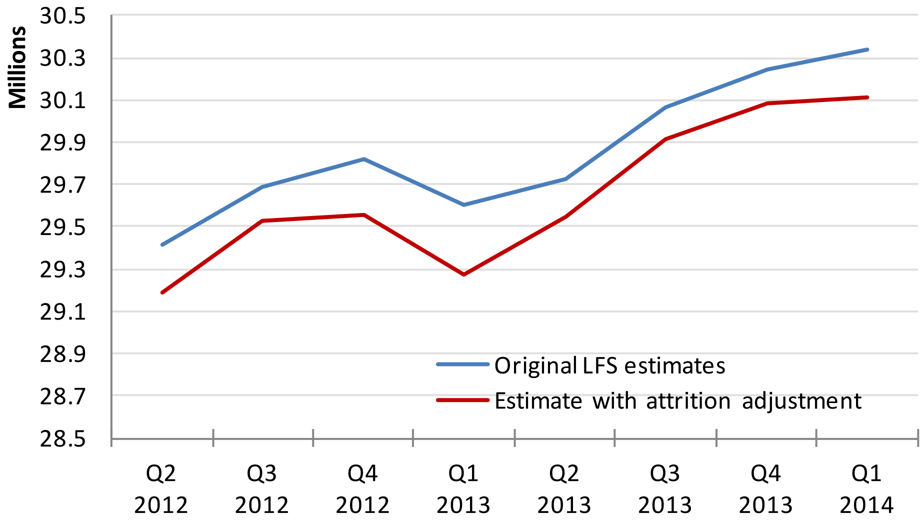
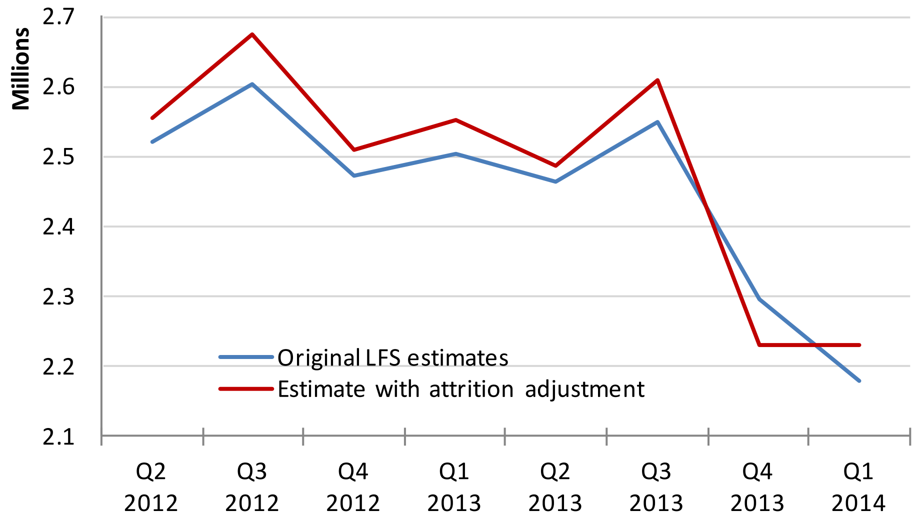


Figure 3b: Total unemployment (aged 16+), when an attrition adjustment is applied.



Estimates for headline Labour Market totals under this new weighting scheme are shown in figures 3a to 3c. LFS estimates are provided for comparison, although it should be noted that these were calculated using ONS 'research' datasets and were not seasonally adjusted, and so will differ from the official published figures.

Figure 3c: Total inactivity (aged 16+), when an attrition adjustment is applied.

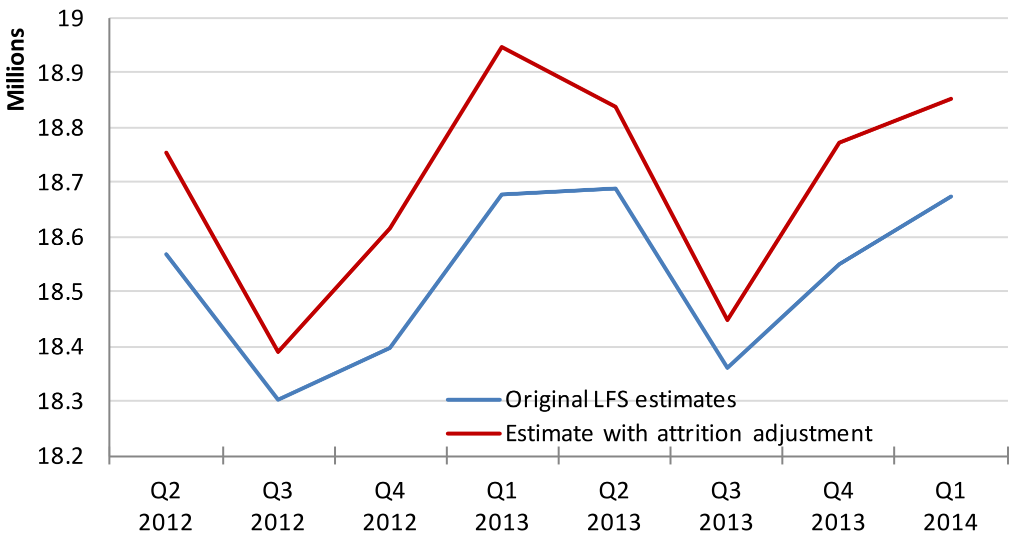


Table 3 contains the average difference between the LFS estimates and adjusted estimates.

Table 3: Average difference between adjusted and original LFS estimates

Estimate	Average Difference
Total Employment	- 223,201
Total Unemployment	+ 46,426
Total Inactivity	+ 194,968

Employment estimates consistently dropped under an attrition adjustment, with inactivity and unemployment rising. For context, the 95% confidence interval for employment estimates is around plus or minus 150,000, meaning the impact on employment is larger than the confidence interval. This analysis suggests that the attrition which is explained by the attrition model does have a notable and fairly consistent impact on headline LFS totals, although it is important to note that the attrition model explains relatively little attrition, and attrition which is unexplained by the model may also have an impact on estimates.

Since the impact appears relatively stable, the effect on estimates of period-on-period change may be minor, although further analysis on a period of relative instability is needed to investigate further.

Running logistic regression as a part of the monthly LFS production system may not be practical. Estimating regression parameters using a single period and applying this across multiple periods produces approximately similar results, but this would not properly reflect changing patterns in attrition over time. We have therefore investigated other methods of mitigating attrition bias.

5. Sample-based weighting based on tenure only

An alternative 'sample based' adjustment which does not required a regression model would be to adjust design weights using a single variable, ensuring that the design-weighted distribution for that variable at waves 2-5 matches the design-weighted distribution at wave 1. These adjusted design weights could then be calibrated in the same fashion as before. Again, this should reduce attrition bias by giving individuals who are more likely to drop out of the survey a larger weight.

The single variable selected should be a good predictor of attrition and also not utilised in the population-based weighting. Based on our analysis, the best candidate variable for this adjustment is tenure; table 2 in section 3 shows that apart from age, sex and region (which are used in the population-based weighting), tenure is the most powerful predictor of attrition. Tenure has other advantages as an adjustment variable – it is already utilised in the LFS longitudinal weighting, and it has an intuitive relationship with attrition, as individuals in rented accommodation tend to move household more often, which can lead to more drop-outs.

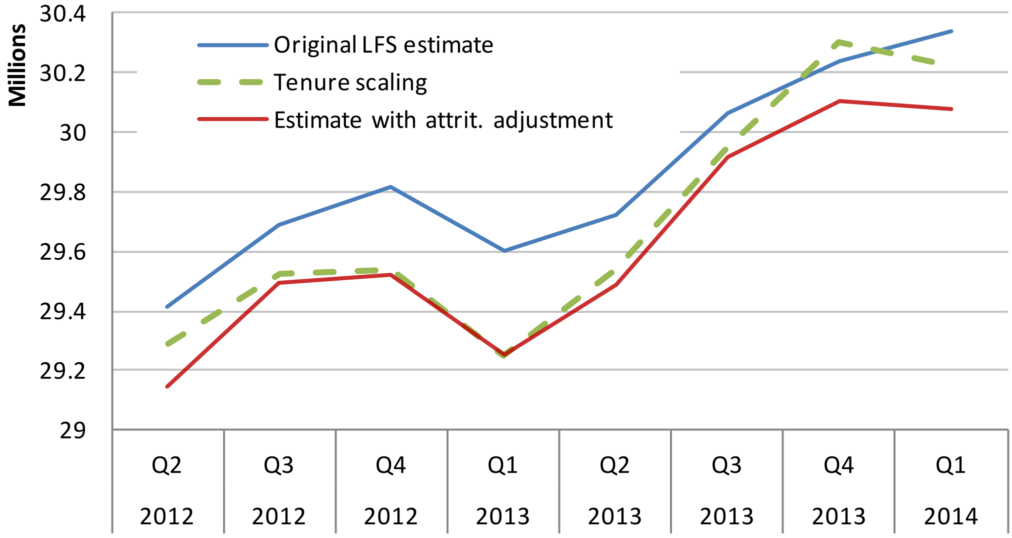
Categories of the tenure variable were combined to create three distinct categories – Owned outright, rented and other. A scaling factor was then produced as:

$$Scaling\ factor\ for\ category\ i, wave\ j = \frac{\frac{Sum\ of\ design\ weights\ in\ category\ i\ at\ wave\ 1}{Sum\ of\ design\ weights\ at\ wave\ 1}}{\frac{Sum\ of\ design\ weights\ in\ category\ i\ at\ wave\ j}{Sum\ of\ design\ weights\ at\ wave\ j}} \quad [1]$$

Employment estimates based on this scheme are compared to unadjusted LFS estimates and LFS estimates calculated using the process outlined in section 4 in the graph below.

Utilising tenure scaling to adjust design weights appears to have a fairly similar impact to the more complex attrition adjustment based on logistic regression, although there is some difference in Q4 2013 and Q1 2014. This may be a more plausible method to adjust attrition bias which could be investigated further.

Figure 4: Total employed (aged 16+), when tenure scaling is applied.

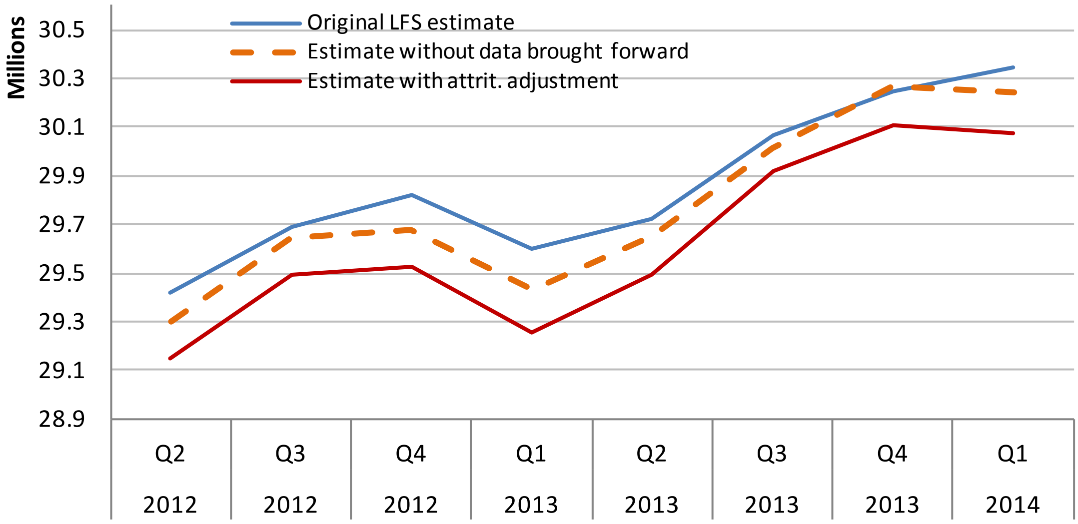


6. Data Brought Forward

One feature of the current survey methodology which may help adjust for attrition is that data for individuals who drop out of the survey through circumstantial refusals or non-contacts (but not data for hard refusals) is rolled forward for one quarter only - referred to as 'data brought forward'. This should in principle reduce attrition bias but may reduce the survey's ability to detect real population change in the short term.

To explore the impact that rolling data forward has on estimates, the data brought forward was removed and estimates were calculated and compared both to the current estimates and to the attrition-adjusted estimates calculated in section 4.

Figure 5: Total employed (aged 16+), when data brought forward is removed.

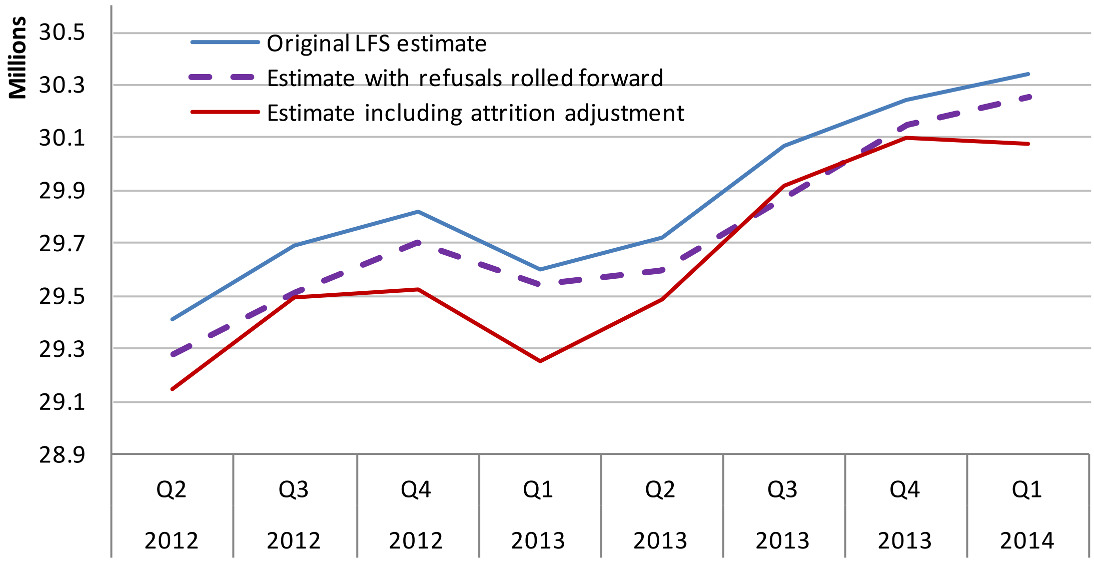


Rolling data forward actually has the opposite effect to adjusting for attrition; the estimates without data-brought forward are closer to the attrition-adjusted estimates. One possible cause of this is the practice for rolling data forwards for only circumstantial refusals and non-contacts, not hard refusals. Figure 5 compares the current estimates to estimates with all hard refusals brought forwards, again including attrition-adjusted estimates.

Rolling forward data for outright refusals in addition to those who are circumstantial refusals appears to bring the estimates closer to the attrition-adjusted estimates, although this is not entirely consistent across all periods.

It is reasonably clear that the practice of rolling data forwards only for circumstantial refusals and non-contacts increases attrition bias. An alternative imputation method, potentially either rolling no data forwards or rolling all data, including for outright refusals, would be preferable. This would need to be subject to further review focused on the impact of any imputation method on the ability of the survey to detect short-term change in the population, since ‘rolling forwards’ can have the effect of ‘smoothing’ real population change in the short-term.

Figure 6: **Total employed (aged 16+), when outright refusals are rolled forward.**



7. Conclusion

Attrition appears to have a notable impact on the levels of headline labour market estimates – applying an attrition adjustment based on a logistic regression model consistently decreases employment by more than the 95% confidence interval, with corresponding increases in inactivity and unemployment. This impact does appear to be fairly consistent, which may suggest that the impact of attrition on short-term estimates of change between periods is limited, although further work on a longer time-span is needed. It is also important to note that our attrition model has relatively low power, and attrition not explained by the model may have an impact on estimates.

We do not recommend implementing a regression-based attrition adjustment in LFS production, as the additional complexity would be challenging to implement and potentially increase the risk of processing errors, and the underlying logistic regression model has relatively low power. Implementing a simpler adjustment to the weighting by scaling the wave-specific design-weighted tenure distribution to match the wave 1 tenure distribution shows some promise as a methodology, and should be investigated further over a longer time series.

The existing method of rolling forwards data for circumstantial refusals and non-contacts but not for hard refusals appears to be increasing attrition bias, and will also reduce the survey's ability to detect short-term change. This method should be reviewed and replaced with an alternative imputation method.

References

Ashworth, K., Merad, S., Weeks, A., and Fallows, A. (2013) "Nonresponse weights for the Labour Force Survey? Results from the Census non-response link study" available at http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/index.html accessed on 1/3/16

Clarke, P. S. and Tate, P. F. (1999) "Production and Analysis of Longitudinal Data from the Labour Force Survey" *GSS Methodology Series no. 17,* available at http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html accessed on 1/3/16

Kanabar, R. (2013) "Accounting for attrition in the 2 quarter UK Labour Force Survey", unpublished paper.

Office for National Statistics (2014) "Review of the Labour Force Survey" *National Statistics Quality Review: Series 2 report 1 available at:* http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/list-of-current-national-statistics-quality-reviews/nsqr-series--2--report-no--1/index.html

Office for National Statistics Labour Force Survey user guidance volume 1 – Background and methodology, available at : https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance/volume1.pdf

Appendix 1

Table A1 – Odds ratios for remaining in the LFS, from logistic regression

Region	Odds ratio	Age band	Odds ratio	Ethnicity	Odds ratio	Household type	Odds ratio	Tenure	Odds ratio	Disability Status	Odds ratio
Tyne and Wear	1.013	0-4	1.438	Item non-response	0.625	1 person	1.116	Item non-response	0.385	Item non-response	0.576
Rest of North East	0.908	5-9	1.501	White	1.234	2 or more persons, all different family units	0.569	Owned outright	1.402	DDA disabled and work-limiting disabled	1.100
Greater Manchester	0.849	10-15	1.595	Mixed /Multiple ethnic groups	1.210	Married couple	1.174	Being bought with mortgage or loan	1.264	DDA disabled	1.062
Merseyside	0.870	16-19	0.647	Indian	0.889	Cohabiting couple	0.934	Part rent	1.353	Work-limiting disabled only	1.443
Rest of North West	0.885	20-24	0.475	Pakistani	1.019	Couple	1.059	Rented	0.911	Not disabled	1.030
South Yorkshire	1.001	25-29	0.598	Bangladeshi	1.326	Lone parent	1.014	Rent free	1.189		
West Yorkshire	1.249	30-34	0.703	Chinese	0.940	2 or more family units	0.927				
Rest of Yorkshire & Humberside	0.947	35-39	0.781	Any other Asian background	1.099	Same sex couple	0.850				
East Midlands	1.046	40-44	0.791	Black /African /Caribbean /Black British	0.988	Civil partners	1.699				
West Midlands Metropolitan County	1.081	45-49	0.867	Other ethnic group	0.875						
Rest of West Midlands	1.070	50-54	0.943								
East of England	1.031	55-59	0.981								
Inner London	0.909	60-64	1.153								
Outer London	0.898	65-69	1.376								
South East	1.063	70 and over	2.861								
South West	1.132										
Wales	0.964										
Strathclyde	0.927										
Rest of Scotland	1.001										
Northern Ireland	1.281										

Using machine learning techniques to clean web scraped price data via cluster analysis

Matthew Mayhew[1] and Gareth Clews

Summary

ONS are investigating the use of prices obtained by automated collection from supermarket websites ('scraping') to compile price indices. These data are mapped into the COICOP classification structure by a machine learning algorithm based on the product name in the raw data. This can lead to misclassification within the products. Our article describes an approach to reduce the number of misclassifications by utilizing a clustering algorithm designed to identify them. We present preliminary results on the impact to the distributions within each of the 35 grocery categories for which this price information was scraped.

1. Introduction

The UK Office for National Statistics (ONS) have published price indices compiled from web scraped price data [1]. These data come from the websites of three leading supermarkets in the UK. Once the price information is collected from the website it is required that it is mapped to the Classification of Individual Consumption according to Purpose (COICOP). This mapping is done so that the prices are classified on a common classification structure rather than the different classification structures the supermarkets use for their own purposes. The approach to this problem is to apply a supervised machine learning technique. This is not infallible and may lead to misclassification within the 35 grocery categories for which the price information is collected.

Calculating indices based on data including misclassifications can cause bias should the prices of misclassified items evolve differently to the prices of the correctly classified products in the class. Reducing this error is clearly an important practice. This paper describes the use of a clustering algorithm as an initial attempt at this task and shows the resulting changes to the price distributions when applied to the grocery dataset. The intention of the program is to find clusters of prices which did not exist in the initial month for which the data was scraped (June 2014). A major assumption in this application is that there are little to no errors in the collection of price quotes within this month. There are three types of price clusters that the program would thus detect as anomalous; a substantial price change of an item over an extended period of time, the introduction of a new product as well as erroneous measurement of the price (for example if an item at £0.93 were recorded as £93).

[1] matthew.mayhew@ons.gov.uk

2. Method

The size of the dataset (several million rows of data), prohibits a manual search through the data to check for anomalies and misclassifications which could change the results of any indices created. Hence, there is a need for an automatic procedure to perform this cleaning. It was thought that clustering may be a suitable method for doing this.

Clustering of the prices was performed using a density based spatial clustering of applications with noise (DBSCAN) approach. This groups together prices which are close to each other whilst marking as outliers points for which the nearest neighbouring price is sufficiently far away. This algorithm is heavily favoured in scientific literature and is one of the most common data mining algorithms currently employed. It is not, however, without its disadvantages.

Should our price information have a large difference in densities of the regions then the DBSCAN method would not be particularly effective in identifying outliers and the choice of the distance between a point and its nearest neighbour that a price must exceed to be considered an outlier must be chosen appropriately.

2.1. Choice of clustering algorithm

There are four types of clustering method.

1. Hierarchical based clustering
2. Centroid based clustering
3. Model based clustering
4. Density based clustering

We shall proceed to briefly describe these, respectively, to explain our choice of DBSCAN. It is assumed within our data that prices for the same product are homogeneous, whereas the prices across different product groups are heterogeneous. Clustering is sensitive to heterogeneity so it would detect misclassifications where it is caused by varying natural price levels of the products included. Clustering is also an unsupervised machine learning method so it doesn't need a set of labels to work. The choice of the clustering regime is key as each one has differing advantages and disadvantages.

2.1.1 Hierarchical clustering

This is a general family of clustering algorithms which build trees of nested clusters by merging or splitting (depending on whether you take a 'top-down' or 'bottom-up' approach to the method) clusters. The root of the tree has the entire sample of prices, which in our instance is the prices for one of the product categories and we would have 35 trees, and the leaves of the tree are samples of size one, the individual prices.

It is required to know how many clusters we will be classifying the data into as a parameter for the method, or make an a priori decision whether clusters are too far apart to be merged. For the web scraped price data the distributions of the prices are unclear, prone to change and can become multimodal/unimodal from initially being unimodal/multimodal respectively. This means that the choice of the number of clusters for our data is not determined prior to the calculations so hierarchical clustering would be a poor choice for our application.

2.1.2 Centroid clustering

Similarly to hierarchical clustering, centroid clustering takes a predetermined number of clusters and seeks to fit this number of clusters to the data. First a set of means is provided, the centroids of the clusters, and areas of equal variance are drawn around these points. For similar reasons as above this is not best suited to calculating clusters for prices from the web scraped data. Also this method is sensitive to choice of initial centroids, so a different choice may lead to different clusters, which would not be useful in our situation, as we require the same products to be removed on each running of the algorithm.

2.1.3 Model based clustering

Model based clustering assumes that the data comes from a finite mixture model described as follows:

$$f(\mathbf{x};\pi,\theta)=\sum_{j=1}^{c}\pi_j g_j(\mathbf{x};\theta) \qquad [1]$$

where π_j are the mixing parameters such that $\sum_{j=1}^{c}\pi_j = 1$ and the g_j are the probability density functions of individual products/collection of products. Using this finite mixture model, observations $\mathbf{x}_i$ are assigned to clusters depending of the maximum value of the posterior probability:

$$\hat{P}(cluster_j \mid \mathbf{x}_i)=\frac{\hat{\pi}_j g_j(\mathbf{x}_i;\hat{\theta})}{f(\mathbf{x}_i;\pi,\theta)} \qquad [2]$$

For this to work the parameters $(\boldsymbol{\pi},\boldsymbol{\theta})$ have to be estimated either via Maximum Likelihood methods or the Expectation-Maximisation algorithm, but these rely on an having a parametric model to start with. However, prices for grocery products are not well distributed and follow no general model[2], as they are frequently multimodel. This means that this approach is not suited to our application.

2.1.4 Density based clustering

Density based clustering requires only a distance and the number of points which must be within this segment for the end points to be considered part of the same cluster. This makes it suitable for use on our dataset. Here, a cluster is a subset of the data with a higher density than the average density across the whole dataset. Density is measured as the number of points within a volume unit; this volume unit differs depending on the algorithm used. The higher the number of points inside the volume unit the more density it has.

There are various density based clustering algorithms. DBSCAN is more common and more highly optimised than almost any other so lends itself well to our needs, due to its relative simplicity among the other algorithms. The DBSCAN algorithm uses a hyperball of radius *r* as its volume unit; this is called a *r*-neighbourhood. DBSCAN depends on two input parameters, *r,* the radius of the hyperball and *m*, the minimum number of points, which in effect choose the minimum density cluster to be found. The algorithm works as follows:

1. Choose a value for *r* and a value *m*
2. Select a point *p*
3. Find the set S_p of all points *q* such that $\| p - q \| \leq r$
4. If the number of points in S_p greater than *m* then label the point a core point, else label the point a border point.
5. Choose another point *p* and repeat steps 2-4
6. If a point *q* is in a set S_p and *p* is a core point then *q* can be directly reached from *p*, label point as directly reach
7. If *p* can be reached from *q* via a set of points $\{p_1,p_2,\ldots\}$, then label reach
8. If two points *p,q* can be reached from another point *t*, then label connected.
9. If a point is labelled reach and connected then label cluster, else label noise
10. Repeat steps 1-9 until all points are labelled either cluster or noise.

For a more thorough derivation and explanation please see Ester *et al.* [3].

In order to clean the data we have made an assumption that there are no misclassifications or other errors during the first month of the collection. This is aided by the fact that during the development of the web scrapers the first month's data was checked thoroughly for these problems and any misclassifications changed or removed as required.

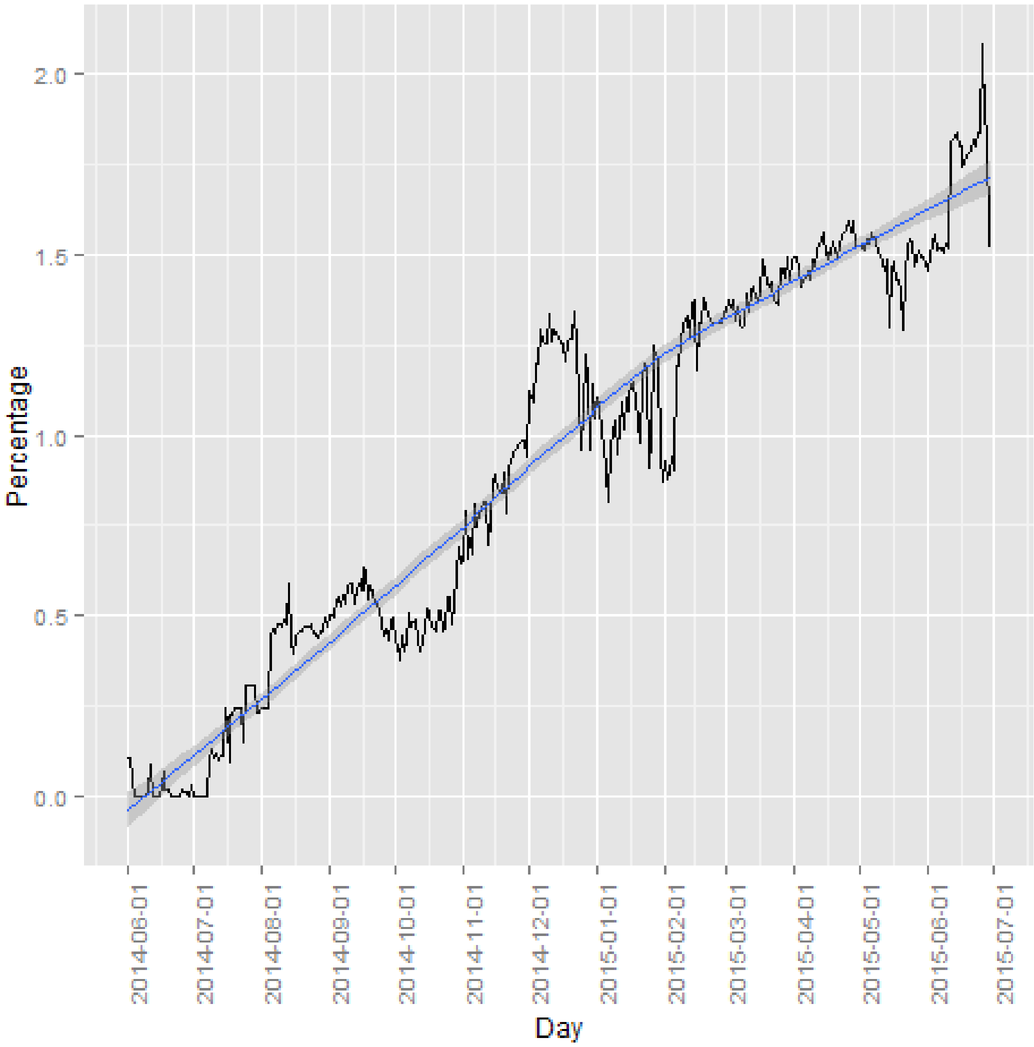
First, clusters are calculated for the data originating in the first month. Next, clusters are calculated for the entire dataset and matched to the clusters obtained for the first month. Prices are tracked across time but products retain the value of the cluster allocated to them within the first collection. This allows us to track the movements of points within clusters and remove those for which there are large deviations in price.

3. Results

3.1. The effect of cleaning

It is clear that the size of the set of items matched to those available in the first period is decreasing over time. This is due to product churn. That is, the turnover of item availability as new items become available, old items disappear and items may drop out temporarily as they are not available in the local supermarket before being reintroduced[2]. Figure 1 shows the percentage of items removed by day. This incorporates both the items not available in the first period as well as the observations removed by the algorithm where prices are available. As a proportion of the whole datasets, the cleaning removes 1% of the products, though as seen in figure 1, more items are removed as a proportion of the items scrapped that day the further away from the training period the day is.

Figure 1. Percentage of items removed from the dataset per day owing to the fact that they are not present on the initial date, blue line shows the estimated trend in removals.



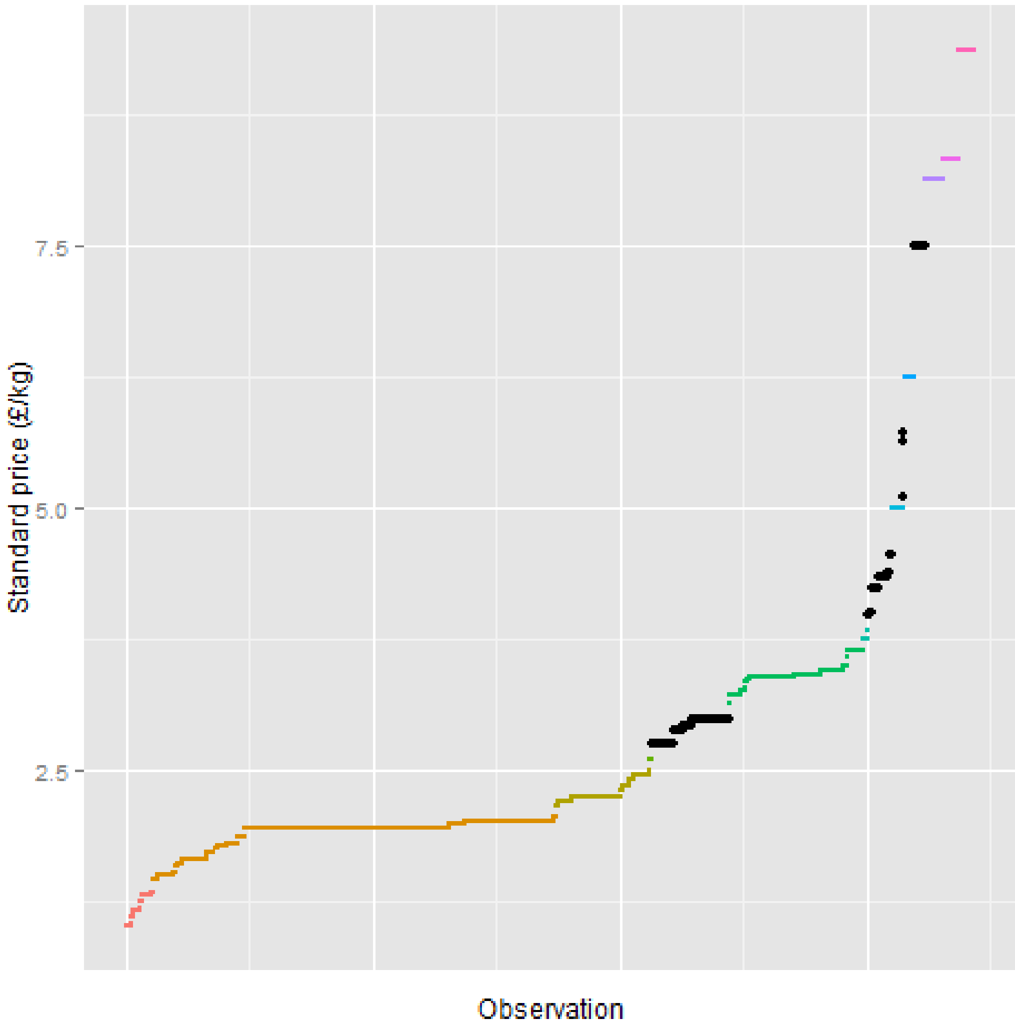
It is of interest to see roughly how many prices are removed for each category and where within the distribution of prices for each category the removals occur. Figure 2 shows the price of apples obtained through the web scraping project during the period June 2014 to June 2015 inclusively. The clusters are colour coded to indicate the grouping of the values. The observations which are identified as erroneous are

[2] Each retailer's website obtains your geographic location and only presents you with the products available in your local stores.

identified in larger point and coloured black. There is a large region near the centre of the distribution at around £2.50 per kg which is removed.

In total, 2357 price points from a total of over 17,000 prices are removed within this category across the 13 months. This includes 24 unique product identifiers which do include several different types of apple. However, the clustering algorithm also identifies and removes several varieties of pear and rhubarb. It is believed that the apple price quotes being removed are due to a price for these specific species of apple not being available in the first month. Figure 1 does not show the split of the removal of products not available in the first period and those removed as misclassifications.

Figure 2. Prices as collected for apples. The larger, black points have been removed by the cleaning algorithm.



To confirm this, the products which were removed by the algorithm were checked to see if they were misclassifications or products not available in the first month. The results are shown in Table 1. There are several areas which immediately draw attention. All of the removals within Brandy are in fact Courvoisier, which does not, as cognac, meet the ONS definition of the Brandy item in the CPI. Whereas all of the removals in the Grape categories are due to new products.

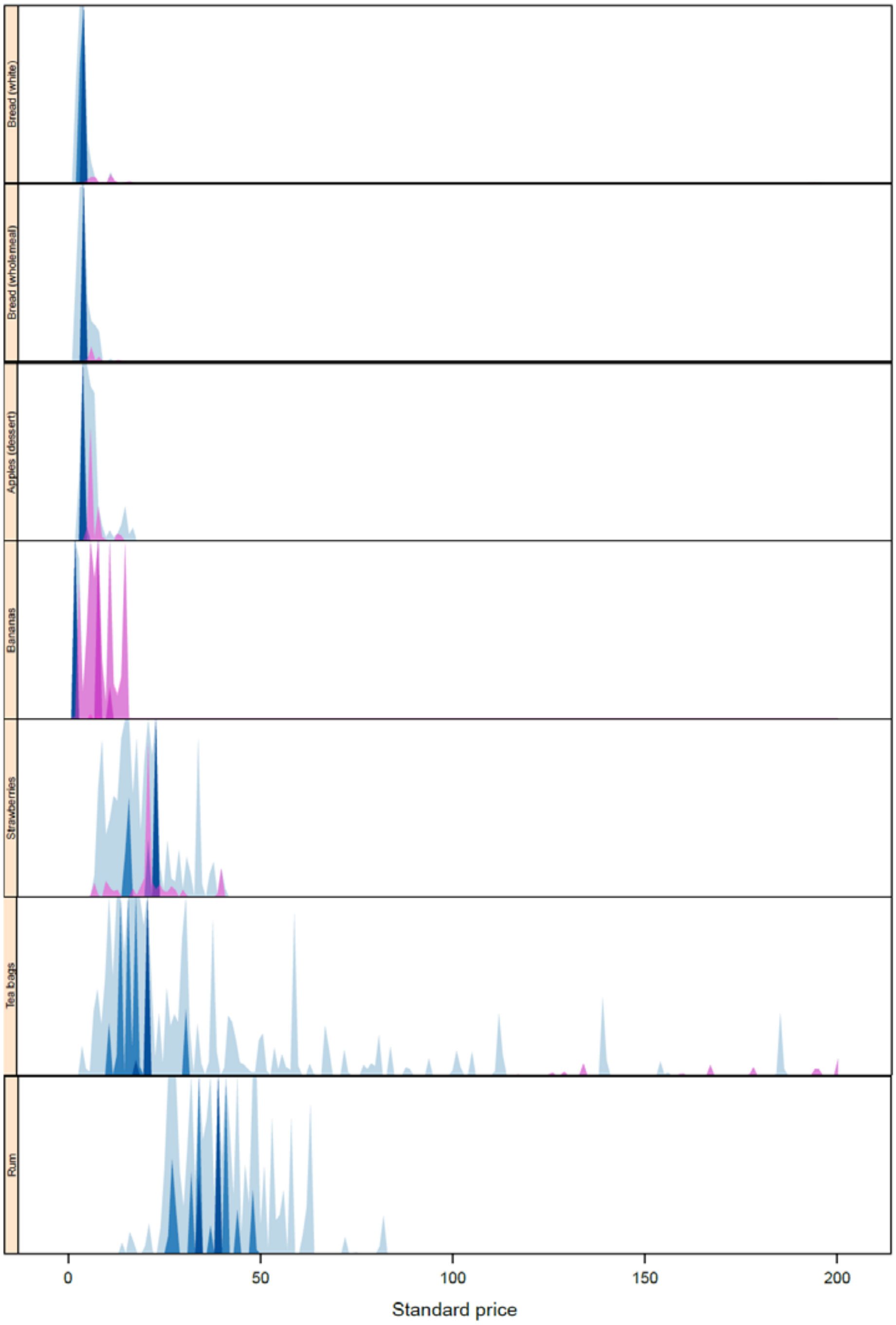
Table 1. Proportion of removals within categories which are misclassifications

Category	Percentage of misclassifications
White bread	0
Wholemeal bread	18
Pasta	28
Cereal	0
Biscuits	73
Whole Milk	100
Yoghurts	75
Tea bags	0
Cola drinks	100
Orange juice	100
Potatoes (not new)	53
Tomatoes	12
Onions	36
Apples	22
Bananas	95
Strawberries	68
Grapes	0
Brandy	100

As an extension of this example considering apples, we calculate kernel density estimators for the prices of each of the 35 classes of grocery product within the data. Figure 3 shows these densities in a horizon plot[4] for seven items. Overlaid onto these are the kernel density estimators for the removed price distributions. A kernel density estimator is a non-parametric method to estimate the probability density function. A kernel is a non-negative function which integrates to one and has expectation equal to 0. The choice of this kernel does have an effect on the visual aspect of the plot, but the shape is the same. The kernel chosen for figure 3 what the Gaussian Kernel.

What is interesting is those cases where the removals have high density and are restricted to one region of the larger distribution of all of the prices within the class. In this instance a geometric mean of the prices would be biased in the direction of the centre of the distribution of misclassified prices. If the location of this distribution changes over time then there would be an impact on any indices calculated from it. It is important to understand the limitations of any automated process for classification of prices and the applications to which the prices are put and this is very much one of them.

Figure 3. **Horizon Plot of the density of cleaned (blue) and removed (pink)**

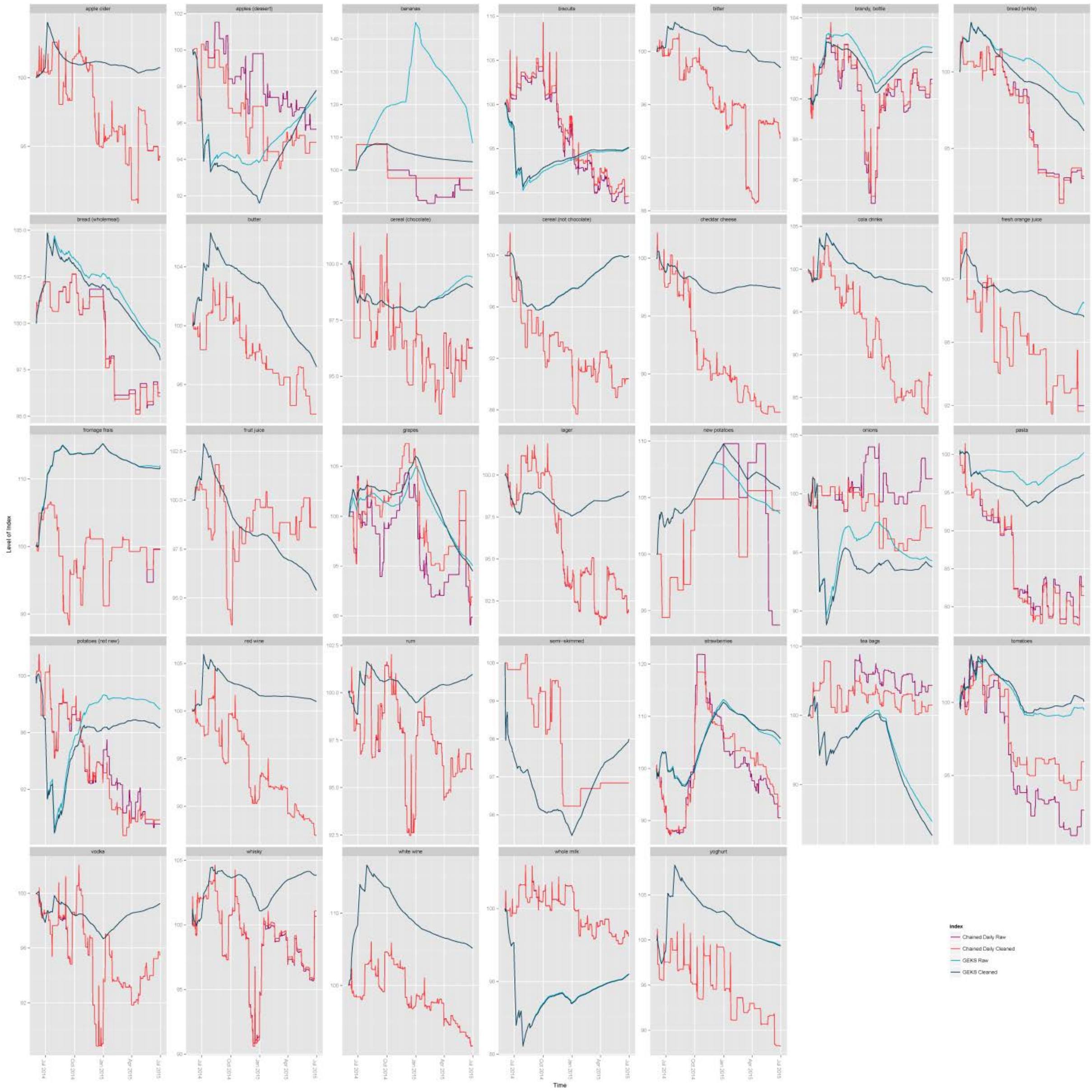


3.2. Effect on Price Indices

From cleaning the prices, it was then useful to see the effect on the price indices that were created in the previous research. The two index formulae that were chosen were the chained daily index and the GEKS index. Figure 4 shows the time series, cleaned and raw, for each formula.

It can be seen that for the majority of the categories there is a noticeable difference between the cleaned and raw data. The most noticeable of these is in the bananas. The cause of the large difference is that around January 2015 one of the websites changed its classification system and placed bananas and grapes in the same category and our web scrapers collected this information. Referring back to figure 3, the removed density is at the more expensive items, so there is an upward bias in the index which is shown in figure 4. Tomatoes are an example where there is a downward bias due to the misclassified items. The indices calculated using the cleaned data are higher than with the raw dataset.

Figure 4. The GEKS and Daily Chained indices for the raw and cleaned dataset



It was then decided to compare the results of the cleaning with the Support Vector Machine, SVM, and other machine learning techniques which were used to classify the data, to find instances when they agree and disagree with each other. In doing this the efficacy of the cleaning could be found as this creates smaller subsets in each category which have the following labels:

- Agreement in classification – both the clustering and the SVM agree in that the product belongs to that category
- Agreement in misclassification – both the clustering and the SVM agree that the product does not belong to that category
- New Product – the clustering says the product does not belong to that category but the SVM says it does, possibly due to the product not being available in the training period
- Disagreement in classification – the clustering says that the product does belong to that category but the SVM does not.

Table 2 Example of Manual Checking for the Bread (White) Category

Product Name	Comparison	Manual Check
Hovis soft white bread, doorstep 800g	Agreement in Classification	Correct Agreement in Classification
Thick sliced soft white Farmhouse loaf 800g	Agreement in Misclassification	Incorrect Agreement in Misclassification
Kingsmill crustaway white bread	New Product	New Product Correct
Warburtons Blackpool milk roll 400g	Disagreement in Classification	Clustering Incorrect
White sliced bread	Disagreement in Classification	Machine Learning Incorrect

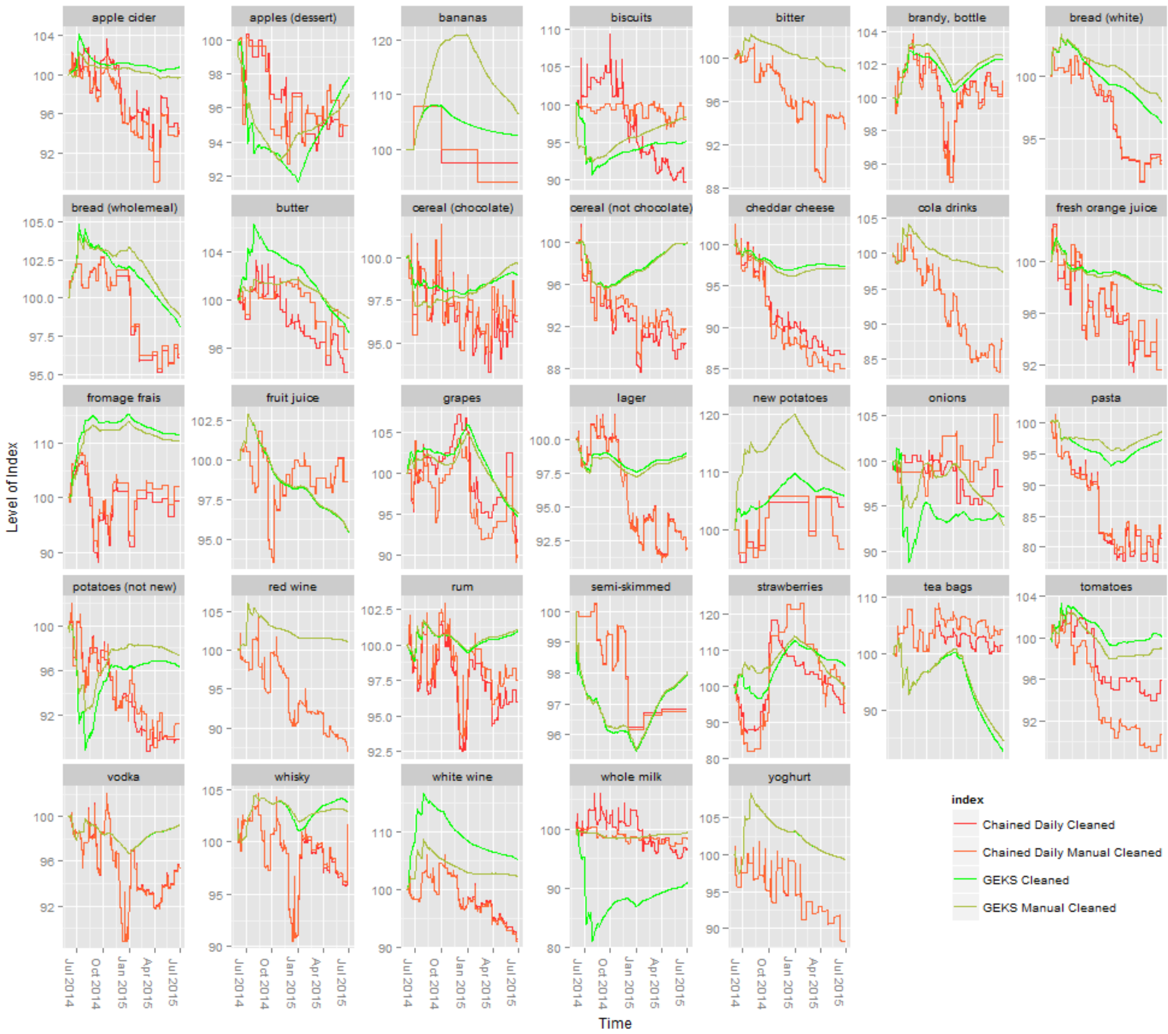
This creates smaller datasets in each of the categories and therefore the set of unique product names in these datasets means that a manual check is needed to see if they are correctly in the subsets or not. Table 2 shows an example of this check for the bread (white) category.

Out of these products only four out of the five should be in the bread (white) category, “Warburtons Blackpool milk roll 400g” does not fit within the stringent rules that the ONS use for this category. The reason why “thick sliced soft white Farmhouse loaf 800g” is labelled Agreement in Misclassification is that there was a 100p error when collecting the price from the website. A new dataset was then created by keeping all the products labelled with “Machine Learning Incorrect”, “Correct Agreement in Classification”, “Incorrect Agreement in Misclassification” and “Correct New Product ”. This Manual Clean dataset was then used to calculate the indices.

Figure 5. shows the effect of the removal of all of the misclassifications. In some categories this only changes the level to a small degree, i.e. they are within one percentage point of each other. There are still some interesting changes when using the manual cleaned dataset. The GEKS for the Whole Milk category shows that the price for whole milk does not change that much, and on just plotting this index the pattern of price movements follows the same pattern as the nearest similar category, Semi-skimmed milk. The misclassifications for whole milk are those products that are in the same aisle as whole milk in the physical shop, products such as milkshakes, this shows that the supermarkets websites are set out similarly.

However are these differences statistically significant? To test this, the differences between the raw and the cleaned, the raw and the manual cleaned, and the cleaned and the manual cleaned were calculated, and then an appropriate hypothesis test was performed. A Mann-Whitney test was used as the data was not normally distributed. Figure 6 shows the p-value of the different hypothesis tests.

Figure 5. The GEKS and Daily Chained indices for the raw and cleaned dataset



For the daily indices all except four of them are significant, this usually corresponds to when there is a visible difference on the graphs for these indices. Table 3 shows proportion of significant differences for each frequency.

Notice that as the frequency decreases the proportion of significant differences decreases, this might be due to the amount of data used to calculate the index, which is different due to the methodology behind the two indices, or it could be that the distribution of the differences becomes more skewed and therefore a Mann-Whitney test may not be theoretically correct to use.

Figure 6. **p-values for the Mann-Whitney test on the differences between the daily indices produced from the cleaning regimes. The red shades are the Chained Daily and the green shades are the GEKS**

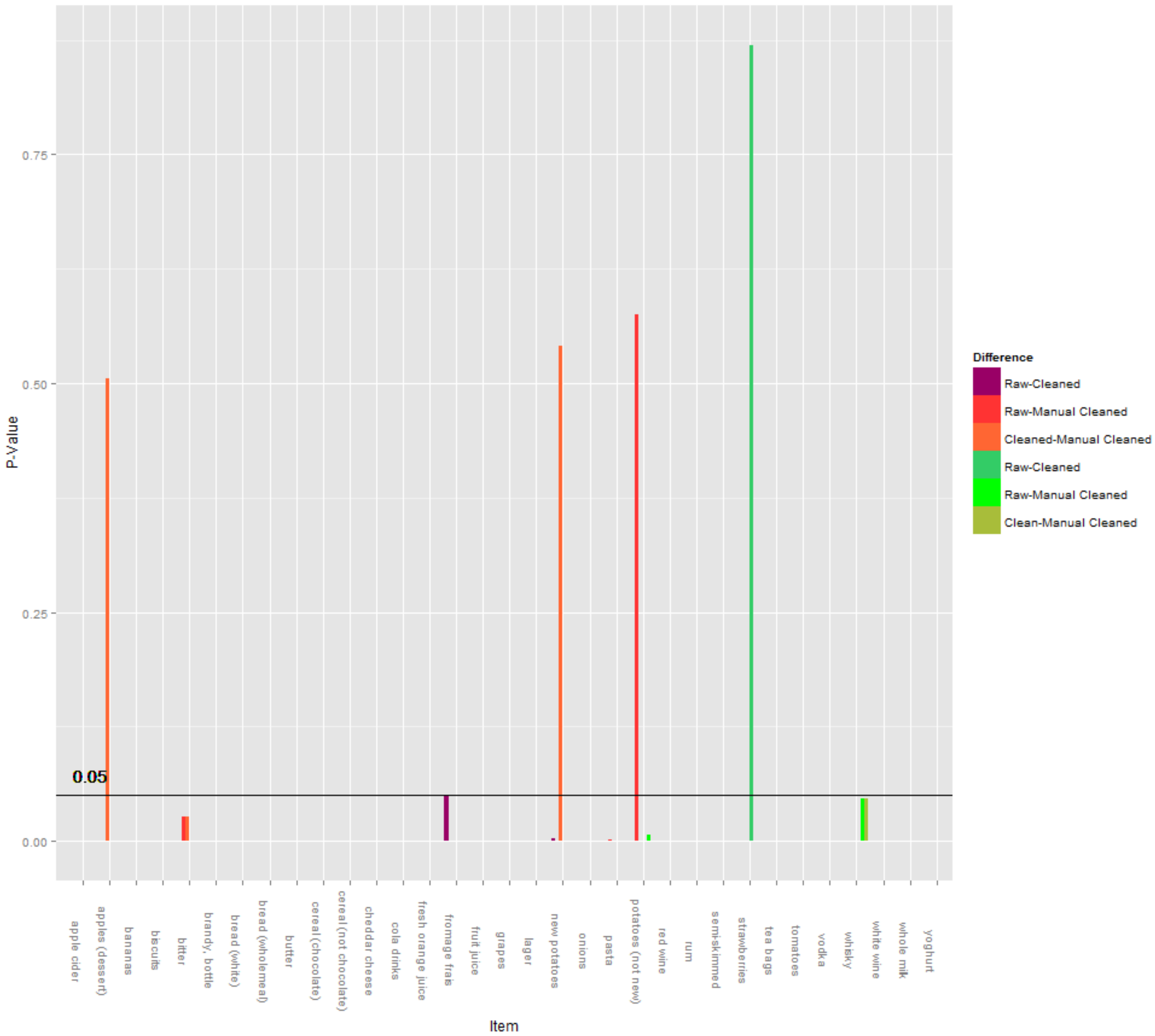


Table 3. Proportions of significant differences.

			Chained			GEKS
Frequency	Raw - Cleaned	Raw - Manual Cleaned	Cleaned - Manual Cleaned	Raw - Cleaned	Raw – Manual Cleaned	Cleaned - Manual Cleaned
Daily	1.00	0.97	0.94	0.97	1.00	1.00
Weekly	0.94	1.00	0.97	0.88	1.00	0.85
Fortnightly	0.88	0.85	0.88	0.94	0.70	0.82
Monthly	0.70	0.67	0.67	0.88	0.70	0.79

3.3. Predicting the Manual Cleaned Labels

Since the Manual Cleaned dataset produces significant differences, these labels need to be predicted as it may not be practical to do the Manual Check each time new prices are obtained. Two approaches were looked at to predict these labels, these were:

- Principal Component Analysis
- Naïve Bayes Classification

3.3.1 Principle Component Analysis

Principal Component Analysis (PCA) is a method that reduces the dimensionality of the dataset while accounting for as much of the original variation as possible. This is achieved by transforming the variables $X_1,\ldots,X_p$ to a new set of uncorrelated variables $T_1,\ldots,T_p$ which are linear combinations of the original variables. The new variables are ordered in a way such that the first few $T_1 \ldots T_d$ with $d<p$ can be selected and will account for most of the variation in all the original variables. These new variables are orthogonal to each other. The orthogonal variables are called scores or components. To find these scores, an optimisation procedure is performed.

First we need to define some notation. Let X be a matrix that contains our data, W be a weights matrix that maps X onto a matrix T which contains the scores. To calculate the columns, $\mathbf{w}_{(\mathbf{k})}$, of the weights matrix W, do the following. For the first column find,

$$\boldsymbol{w}_{(\mathbf{1})} = \underset{w}{\operatorname{argmax}} \frac{\boldsymbol{w}^{\boldsymbol{T}} X^T X \boldsymbol{w}}{\boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{w}} \qquad [3]$$

Once the first column is found an iterative procedure begins with the next two steps until all columns are found. To find the kth weights column, first remove the influence of the k-1 previous columns from X, as follows:

$$\hat{X}_k = X - \sum_{i=1}^{k-1} X \boldsymbol{w}_{(\boldsymbol{i})} \boldsymbol{w}_{(\boldsymbol{i})}^{\boldsymbol{T}} \qquad [4]$$

Then find then the maximum argument of the same function as before

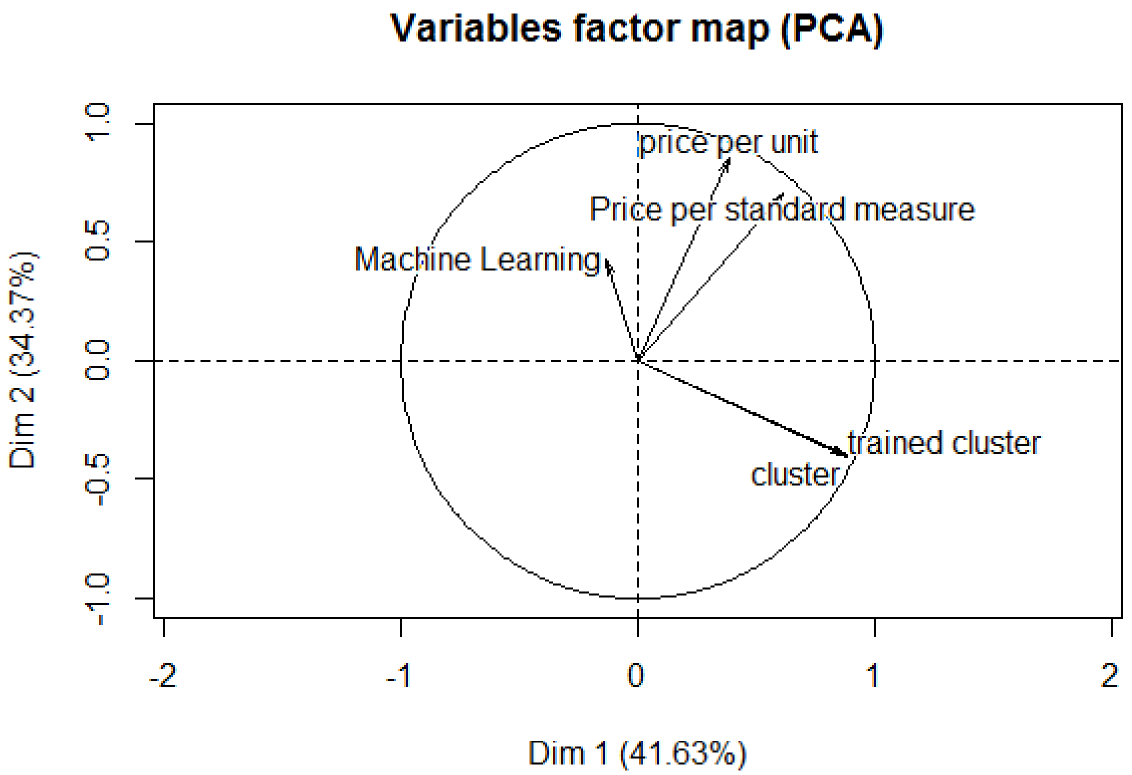
$$\boldsymbol{w}_{(\boldsymbol{k})} = \underset{w}{\operatorname{argmax}} \frac{\boldsymbol{w}^{\boldsymbol{T}} \hat{X}_k^T \hat{X}_k \boldsymbol{w}}{\boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{w}} \qquad [5]$$

Therefore the resulting Scores Matrix is calculated by T=XW.

A PCA was performed on all of the variables in the dataset and then on just a subset as it was found that most of the variables, such as the day the price was collected on and the units the product is measured in did not have much weight in the matrix W.

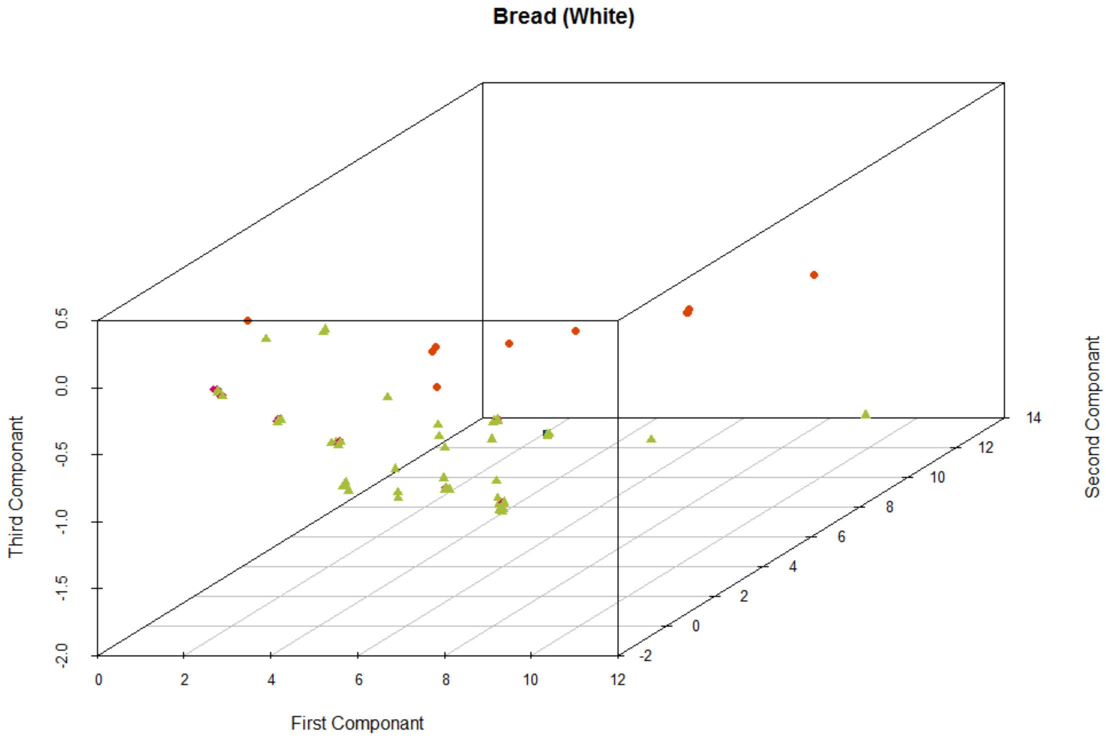
Analysing the result it was found that the first three components cover over 80% of the variation so could be used to predict the labels. To see which of the measured variables had more influence on the scores, a circle of correlations was created, see figure 7. The longer the arrow is on the circle of correlations the more influence it has on the score; also the cosine of the angle between the axis and the arrow is the correlation between the variable and the score [5]. As can be seen the price variables have more influence on the second score and the cluster variables have more influence on the first score.

Figure 7. Circle of Correlations for the first two scores.



Plotting the first three scores for the Bread (white) category and colouring the points according to the label that the manual cleaning gave it, we can see that PCA does not produce an arrangement of the variables such that the different labels are separable, i.e. there is a group of points with one label and another group with another label that a line can be drawn through. Figure 8 shows this as the Green circles and the pink diamonds cannot be split. Therefore another method should be used to predict the labels.

Figure 8. Data points in first three scores, there are four different labels in this category.



3.3.2 Naïve Bayes Classifier

A Naïve Bayes Classifier finds the label, L_i which maximises the conditional probability $P(L_i|x)$ where $x=(x_1,\ldots\ldots,x_n)$ are a realisation of the values the other variables take. Using an application of Bayes Rule and the assumption that the variables are class conditionally independent, the classifier maximises this expression:

$$P(L_i|\boldsymbol{x}) = \frac{P(L_i)\prod_k^n P(x_k|L_i)}{P(\boldsymbol{x})} \qquad [6]$$

Table 4. Sample of Results for Naïve Bayes with additive smoothing parameter equal to 1

Product Name	Category	True Label	Predicted Label
Cherry punnet 225g	Strawberries	Clustering Incorrect	Correct Agreement in Misclassification
Braeburn apple 6 pack 670g	Apples	Correct Agreement in Classification	Incorrect Agreement in Classification
Cherry vine tomatoes, 400g	Tomatoes	Correct Agreement in Classification	Correct New Product
Davidstow cornish classic, mature cheddar 200g	Cheddar Cheese	Correct Agreement in Classification	Correct Agreement in Classification

Often additive smoothing is done so that each of the conditional probabilities in the RHS are non-zero, as this would make the whole expression zero if not done. The normalisation constant is often ignored as it is usually impossible to know the probability of observing that realisation. The conditional probabilities are based on a set of training data, so will change when more data is available, same happens with the class probabilities. An example of the labels predicted is given in table 4.

There are examples where the true label and predicted label agree with each other and there are examples where they do not agree. When the labels disagree with each other we have three situations, which are:

1. True Label and Predicted Label are labels associated with misclassification and should be removed, as with the cherries.
2. True Label is a label associated with correct classifications and the predicted label is associated with removal, as with the apples, or vice versa
3. True Label and Predicted Label are labels associated with correct classification and should be removed, as with the tomatoes.

The disagreement of the labels is a problem that cannot be avoided, as we are maximising probability, however we can measure how different they are, four such measures will be presented. These measures are:

$$Recall = \frac{\#(True\ Positive)}{\#(True\ Positive) + \#(False\ Negative)} \qquad Precision = \frac{\#(True\ Positive)}{\#(True\ Postive) + \#(False\ Positive)}$$

$$Accuracy = \frac{\#(True\ Positive) + \#(True\ Negative)}{\#(True\ Positive) + \#(True\ Negative) + \#(False\ Positive) + \#(False\ Negative)} \qquad F1 - Measure = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Where # means the number of products with that condition. The closer the value of any of the measures is to one the better our classifier is. To test this, the classifier was trained on one sample of the data and then used to predict the labels for 10 different samples of the data and it was found that Accuracy=0.66, Precision=0.69, Recall=0.97 and F1=0.76 so the classifier does not produce many false positives, which is analogous to a small value for a Type 1 error, from hypothesis testing. This is good as it shows that our classifier is not just randomly assigning labels. Because of this we can use a Naïve Bayes Classifier to replace the Manual Check for further data collections.

3.4. Stability and Size of Clusters

One final item to check is to see if the clusters are stable and if the size of the clusters over time remained the same over time. To do this the length of the time period that the training clusters were created on was reduced to one week. Ideally this should by one day but some of the categories have a very small daily sample size that the algorithm would fail. The clusters were then produced for each week of the year and then compared to see if the same clusters exist for a long period and are of the same size. It was found that for this frequency, the clusters are stable throughout time in the majority of cases. This is because the size of the cluster often changes due to the introduction of offers and new products. When this happens clusters tend to merge together as the new price instances tend to fill the gap between clusters. Figure 8. shows an example of this, where if you look at week 48 in 2014 a product is classed as noise but in week 49 it is part of the neighbouring cluster as there is an extra product between the end of the cluster in the previous week and the noise product.

4. Conclusion and further work

Cluster analysis methods allow us to clean and verify data which is empirically difficult to manually validate, providing us with a method to detect anomalies and clean the dataset in a smaller timeframe than would have happened otherwise. The results of this cleaning are that there are fewer misclassification anomalies influencing the indices produced making the different formulae show similar price movements.

The next steps of this project is to incorporate this algorithm into the web scrapers so that the cleaning becomes automatic and for the output to help improve the classification algorithms that place the products into the categories, which in turn would improve the cleaning process in a feedback loop. Further work to undertake would be to; (1) improve the estimates of the parameters into the clustering algorithm; (2) make the training period a rolling period, as it was noticed that the more items were removed the further away from the training period the prices were collected and, (3) complete a more thorough investigation into product churn and its effect on the clusters, and whether it may count for the difference between the index formulae.

References

[1] Breton, R. et al (2015) Research Indices Using Web Scraped Price Data, September 2010. Available at: http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/research-indices-using-web-scraped-price-data/index.html

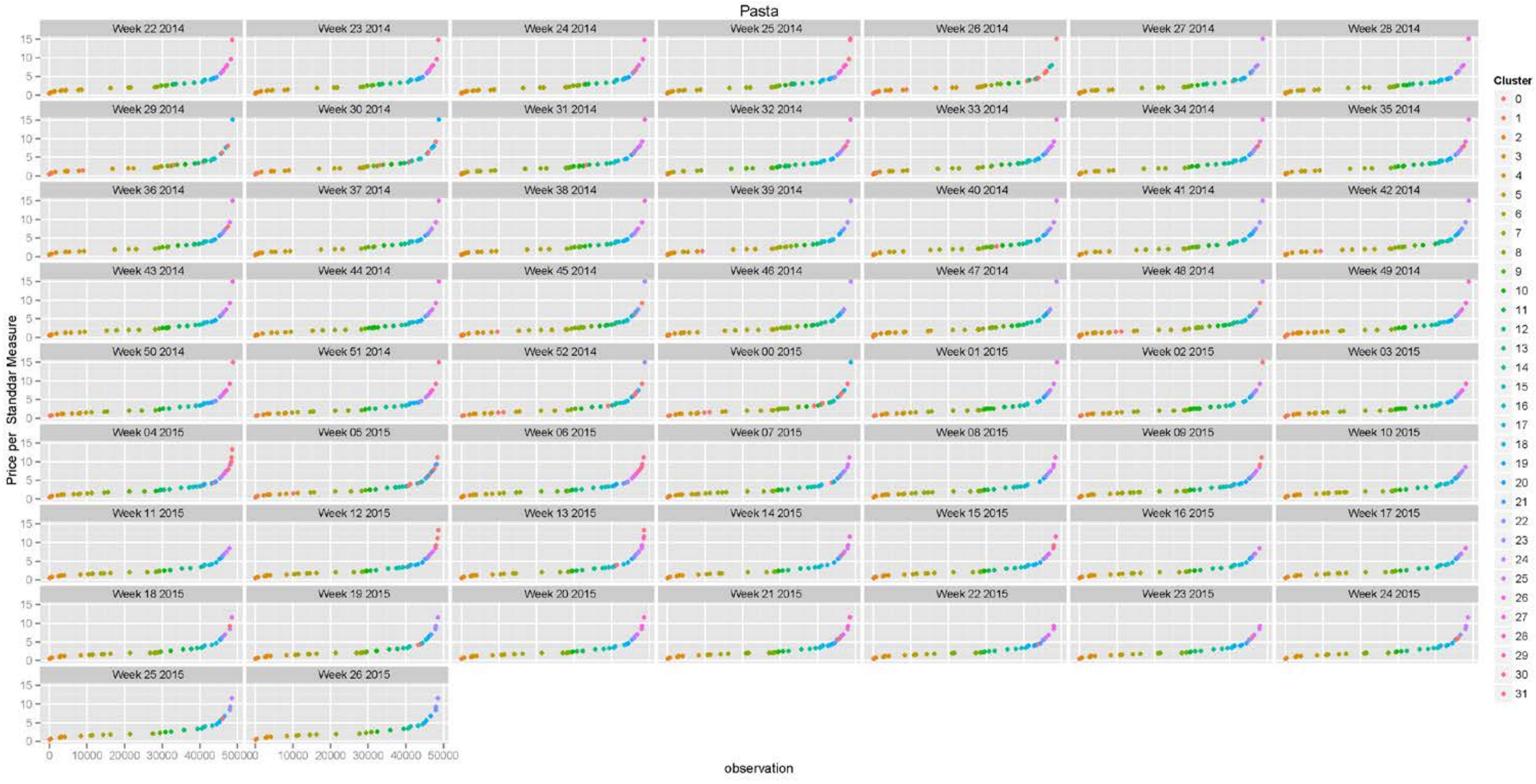
[2] Moulia, N. (2015) An investigation into online vs. offline prices from various shops across the UK, University of Cardiff, Department of Mathematics, MSC thesis, available upon request

[3] Martin Ester *et al.*A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Miningm page 6, 1996*

[3] Heer, J. et al (2009) Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series, University of Berkeley, available at: http://vis.berkeley.edu/papers/horizon/2009-TimeSeries-CHI.pdf

[4] Hervé Abdi and Lynne J. Williams (2010) Principle Component Analysis, WIREs Comp Stat 2010 2 433–459

Figure 8. Pasta clusters for the whole year.



The weighting methodology for Wave Four of the Wealth and Assets Survey

Robynne Davies

Overview

The Wealth and Assets Survey (WAS) is designed to collect information on all aspects of individual and household wealth for private households across Great Britain. WAS began in 2006 and has recently published findings from the Wave 4 (W4) data. Its multi-panel design enables WAS to be used for both longitudinal and cross-sectional analysis.

The WAS longitudinal weighting strategy is based on the principal of adjusting the initial selection probability to compensate for attrition; data collected from previous waves are used to calculate model based-attrition adjustments. By W4 there are many different types of responders whose cross-sectional weights are calculated in different ways. The W4 cross-sectional weighting strategy considers weight-sharing for joiners to households, assignment of weights to births and treatment of responders who leave the sample but later re-join the survey. The ways to adjust for non-response, such as non-response classes and model based non-response adjustments are also considered.

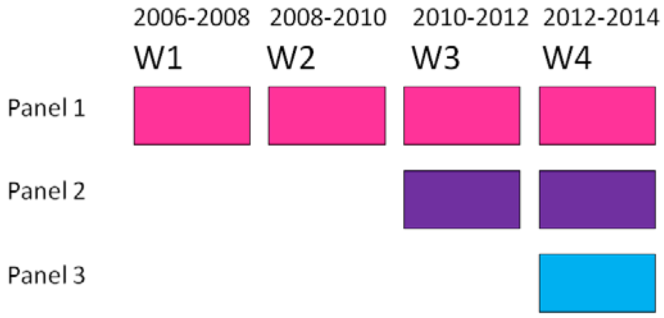
This paper first discusses how the principle of the longitudinal weighting strategy was adapted to compute two different W4 longitudinal weights; the W1-W4 longitudinal weight and the W3-W4 longitudinal weight. It then discusses the many different weighting methods applied when constructing the W4 cross-sectional weight. Finally, the method used to combine multiple panels of weights and the properties of the final weights is discussed.

1. Structure of the Wealth and Assets Survey

WAS began in 2006 with the fieldwork for each wave lasting two years. WAS is designed to follow the same people over time; this is achieved through consecutive waves of interviews. This longitudinal perspective of the survey allows for estimation of gross change over time.

The cross-sectional perspective of the survey is another important feature as it allows for estimation of wealth at points in time and the subsequent estimation of net change over time. Top-up panels have been introduced to WAS in both W3 and W4. Top-up panels boost the sample size which decreases as the survey progresses through the waves, because of attrition in the sample. They also update the sample so that it better represents the population to which the cross-sectional estimates relate. Figure 1 depicts the wave and panel structure of WAS up to and including the fourth wave.

Figure 1 **WAS structure**



Respondents in W1 and W2 of the survey were sampled from panel 1. At W2, individuals who moved into W1 sampled households were also included in the sample but did not remain in the sample if they later left the household. Approximately 30,000 households responded at W1, but attrition saw this reduce to 20,000 households by W2. In order to maintain this sample size, top-up panels were introduced in W3 and W4. Therefore, the W3 sample consisted of respondents from two panels and the W4 sample from three panels.

2. WAS sampling scheme

WAS follows a two-stage stratified cluster design. The primary sampling units (PSUs) are postcode sectors; therefore the first stage is to systematically choose a number of postcode sectors from an ordered list with probability proportional to size. The list of postcode sectors is ordered by region, metropolitan borough status and two census data variables that are correlated with wealth: socio-economic status and proportion of households with no car. A fixed number of addresses are then randomly chosen within each PSU. A file of the addresses in the selected PSUs is sent to HMRC for matching against tax data to identify households likely to be amongst the most “wealthy”. As the distribution for wealth is very skewed, households likely to be in the top 10 per cent of wealthiest households are assigned a higher probability of selection by a factor of 3; this reduces the impact of extreme values on precision and also addresses the fact that it is harder to gain an interview with these households.

The address selection probabilities for addresses from the s^{th} PSU are:

$$\begin{aligned} P(address\ sampled) &= P(PSU\ sampled).P(address\ sampled\ |PSU\ sampled) \\ &= {}^{nN_s}/_{N}\ P(address\ sampled\ |PSU\ sampled) \end{aligned} \qquad [1]$$

where n is the number of sampled PSUs , N_s is the number of addresses in the s^{th} PSU and N is the total number of addresses included on the sampling frame in Great Britain. For an address in the predicted high or low wealth stratum, respectively, the selection probabilities are:

$$
\begin{aligned}
P(high) &= {}^{nN_s}\!/_{N}\; P(address\ sampled\ |PSU\ sampled, address\ is\ high\ wealth) \\
&= {}^{nN_s}\!/_{N} \cdot {}^{3n_{psu\_hi}}\!/_{(M_s^{lo} + 3M_s^{hi})}
\end{aligned}
\quad [2]
$$

$$
\begin{aligned}
P(low) &= {}^{nN_s}\!/_{N}\; P(address\ sampled\ |PSU\ sampled, address\ is\ low\ wealth) \\
&= {}^{nN_s}\!/_{N} \cdot {}^{n_{psu\_low}}\!/_{(M_s^{lo} + 3M_s^{hi})}
\end{aligned}
\quad [3]
$$

where n_{psu_hi} is the number of “wealthy” addresses selected from the selected PSU, n_{psu_low} is the number of “non-wealthy” addresses and M_s^{lo} and M_s^{hi} are the number of addresses in the low and high strata respectively. The design weights for cases selected in panel 3 for the sampled addresses, d_i^{p3}, are then the reciprocal of the appropriate address selection probability.

Table 1 details the number of units sampled at each stage for the three panels, where n_{psu} is the number of postcode sectors selected in each panel and n_{totadd} is the total number of addresses selected from all PSUs in each panel.

Table 1 Number of sampling units in each panel

	n_{psu}	n_{totadd}
Panel 1	2,400	62,400
Panel 2	648	12,636
Panel 3	636	8,268

The PSU selection for each panel was carried out independently. Therefore, by chance, the same PSU may have been selected in more than one panel.

3. Longitudinal Weighting Strategy

The WAS longitudinal weighting strategy is based on the principle of adjusting the initial selection probability to compensate for attrition; that is to adjust for those who drop out of the survey over time. This is achieved through the development of the longitudinal base weight (see e.g. Verma et al. 2007). This principle enables the weights to refer back to the desired population as closely as possible, taking into account the sample design and respondent follow-up procedures.

For W4, the two longitudinal weights viewed as most valuable for longitudinal analysis are produced:

- W1-W4 longitudinal or "survivors" weight; this is calculated for those who respond in every wave up to and including the fourth wave

- W3-W4 longitudinal weight; this is calculated for those who respond in both W3 and W4

They are discussed separately below.

3.1. W1-W4 "survivors" longitudinal weight

As discussed above, the longitudinal weighting strategy is based on adjusting the initial selection probability to compensate for those who drop out of the survey. Therefore, the first step when calculating the W1-W4 longitudinal weight is to choose an appropriate weight associated with the initial selection probability. The W1-W3 longitudinal weight is an appropriate choice as the W1-W4 longitudinal sample is a continuation of the W1-W3 longitudinal sample. The W1-W3 weight would have been calculated as part of the W3 WAS weighting.

After choosing an appropriate weight, an attrition model needs to be developed in order to adjust the W1-W3 weight for those who left the sample between W3 and W4. An adjustment needs to be made for attrition as it a source of potential bias, particularly if those who leave the survey are different from those who remain. WAS attrition is made up of two components:

- The eligibility status of an individual becomes unknown between W3 and W4 (for example, it is unclear whether they have left the country or not, and so they may now be ineligible for WAS)
- Non-response/non-contact at W4

Models for each component of attrition are built separately. First, we consider the model for unknown eligibility status. Logistic regression is used to predict the probability of a case's eligibility status remaining known between W3 and W4 using a wealth of information from W3. Models for both components of attrition are fixed following research carried out in earlier waves of the survey. Variables used in the eligibility model include the accommodation type of a household, the number of children in a household, the length of the previous interview and the ethnicity and age of the respondent. W4-known eligibility cases that have a low probability of remaining known are given a larger weight. This way, cases with similar characteristics to the drop-outs compensate for those who leave the survey.

This can be also be described using the following notation. The logistic regression model gives a predicted probability that the eligibility status at W4 is known for each case, denoted as $\hat{\theta}_i^o$.

The unknown eligibility status weights w_{4i}^{uo}, are then calculated as:

$$w_{4i}^{uo} = {}^{1}\!/_{\hat{\theta}_i^o} \qquad\qquad i \in s_4^o \qquad\qquad [4]$$

where s_4^o is the sample of people with a known eligibility status at W4.

A similar method is used for the non-response adjustment. A second logistic regression model is built using W3 variables to predict the probability of a case responding in W4. Variables used in the non-response model include accommodation type, tenure, whether the respondent has a current account and the length of the previous interview. Responding cases with W3 characteristics that indicate they are unlikely to respond are given higher weights.

Again, this can be described using the following notation. A second logistic regression model calculates the predicted probability that a case would respond in W4, denoted as $\hat{\theta}_i^r$.

The non-response weights w_{4i}^{nr} are then calculated as:

$$w_{4i}^{nr} = {}^{1}\!/_{\hat{\theta}_i^r} \qquad\qquad i \in s_{1-4}^r \qquad\qquad [5]$$

where s_{1-4}^r are W1-4 cases that responded in W4.

For individuals in an eligible respondent household, s_{1-4}^r, the W1-W4 longitudinal base weight w_{4i}^{1-4} is the product of the relevant W3 weight and the two attrition weights i.e. the W1-W3 longitudinal weight $w_{3i_{cal}}^{1-3}$ multiplied by the unknown eligibility and non-response weights. For ineligible individuals known to be part of movements out of the target population at W4, out_{1-4}, the longitudinal weight is the product of the W1-W3 longitudinal weight and the unknown eligibility status weight.

$$w_{4i}^{1-4} = \begin{cases} w_{3i_{cal}}^{1-3} w_{4i}^{uo}\ w_{4i}^{nr} & i \in s_{1-4}^r \\ w_{3i_{cal}}^{1-3} w_{4i}^{uo} & i \in out_{1-4} \end{cases} \qquad\qquad [6]$$

Equation (6) shows that individuals within households that become ineligible are assigned a longitudinal weight. Our true population of interest for W1-W4 longitudinal weighting is the W1 population minus those who become ineligible for WAS by W4. However, we have no known population totals for this population of interest; we only have the W1 time point population totals. Therefore, as a compromise, those who become ineligible are also given a weight. Ineligible cases account for approximately 5 per cent of the W1-W4 longitudinal dataset.

The two longitudinal subsamples (eligible respondents and ineligible outflows) are, after adjustment for attrition, representative of the W1 time-point population. It is therefore possible to calibrate the base weight to the relevant population totals.

The calibration factors, g_i^{1-4} , are calculated to minimise the distance between the pre-calibration weight and the calibrated weight while summing to a set of known calibration totals. WAS calibration groups are:

- 24 groups of sex by age (e.g. males aged 16-24)
- 11 regional groups (e.g. the North East)

The population totals used are based on ONS' mid-year estimates taken from the midpoint of the W1 fieldwork period. As W1 fieldwork took place July 2006 – June 2008, the June 2007 estimates are used (see figure 1).

The final W1-W4 longitudinal weight is the product of the relevant W3 weight, attrition adjustments and calibration adjustment.

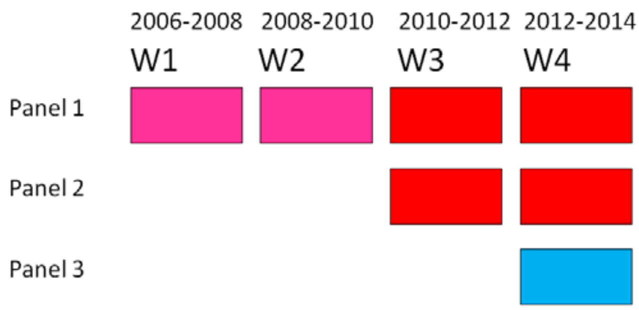
$$w_{4i\_cal}^{1-4} = \begin{cases} w_{3i_{cal}}^{1-3} w_{4i}^{uo} \; w_{4i}^{nr} \; g_i^{1-4} & i \in s_{1-4}^r \\ w_{3i_{cal}}^{1-3} w_{4i}^{uo} \; g_i^{1-4} & i \in out_4 \end{cases} \qquad [7]$$

3.2. W3-W4 longitudinal weight

The W3-W4 longitudinal weight follows the same weighting method as described in Section 3.1, with a few subtle differences and added complexities which are discussed below.

As mentioned in Section 1, a top-up panel was introduced to the survey in W3. As a result, a W3-W4 longitudinal responder can either have been sampled from the original panel or the W3 top-up panel as demonstrated below in figure 2. Here, the red boxes combine to give the W3-W4 longitudinal sample. As the two panels were sampled at different time points, it is more appropriate to weight the two panels separately and combine at the end.

Figure 2 **W3-W4 multiple panels**



To recapitulate the method outlined in section 3.1, the WAS longitudinal weighting method involves choosing a weight associated with the initial selection probability, adjusting this for attrition and finally calibrating to known population totals. The W3 cross-sectional weight, w_{3i}^{xs}, is chosen as a starting weight for both panels; this is a suitable choice as all W3-W4 longitudinal responders will have a W3 cross-sectional weight by definition. This is then adjusted for unknown eligibility and non-response using the same methods as previously described. Therefore, the W3-W4 longitudinal base w_{4i}^{3-4} for a case belonging in panel k can be defined as:

$$w_{4i}^{3-4} = \begin{cases} w_{3i}^{xs}\, w_{4i^*}^{uo}\, w_{4i^*}^{nr} & i \in s_{3-4}^{r_k} \\ w_{3i}^{xs}\, w_{4i^*}^{uo} & i \in out_{3-4}^{k} \end{cases} \qquad [8]$$

where $k = 1, 2$, $s_{3-4}^{r_k}$ is the set of W3-4 cases that responded in W4 in panel k and out_{3-4}^{k} is the set of ineligible W3-W4 cases in panel k. Asterisks are used in equation (8) to distinguish the W3-W4 attrition adjustments from the W1-W4 attrition adjustments. The W3-W4 longitudinal base weight is then calibrated to the W3 time-point population totals, the June 2011 mid-year estimates.

The final W3-W4 longitudinal weight is the product of the relevant W3 weight, attrition adjustments and calibration adjustment.

$$w_{4i\_cal}^{3-4} = \begin{cases} w_{3i_{cal}}^{1-3} w_{4i^*}^{uo}\ \ w_{4i^*}^{nr}\ g_i^{3-4} & i \in s_{3-4}^{r_k} \\ w_{3i_{cal}}^{1-3}\ w_{4i^*}^{uo}\ \ g_i^{3-4} & i \in out_{3-4}^{k} \end{cases} \qquad [9]$$

3.2.1 Combining the panels

There are two panels that contribute to the W3-W4 longitudinal weight. As each panel has been weighted separately they need to be combined. One way of doing this would be to join the panels together with respect to the achieved sample size of each panel. This takes into account the proportion of cases each panel contributes to the total number of W3-W4 longitudinal cases. This is achieved by multiplying each calibrated weight $w_{4i\_cal}^{3-4}$ by the following factor:

$$factor_{1=}\ {}^{n_k}\!/\!_{\sum_k n_k} \qquad i \in \text{panel } k \qquad [10]$$

where n_k is the number of cases in panel k ($k = 1, 2$ in this case).

An alternative method is to join the panels together with respect to the effective sample size of each panel. The effective sample size is the sample size required under a simple random sampling scheme that will yield the same variance for an estimate that has been produced under the true, more complex design. This is achieved by multiplying each weight by the following factor:

$$factor_2 = {neff_k}\Big/{\sum_k neff_k} \qquad i \in \text{panel } k \qquad [11]$$

where $neff_k$ is the number of cases required under a simple random sample to yield the same variance for an estimate produced under the current design in panel k.

The effective sample size for each panel is calculated using Kish's approximate formula for the effective sample size:

$$Kish\, neff = \frac{(\sum_i w_i)^2}{\sum_i (w_i{}^2)} \qquad i \in \text{panel } k \qquad [12]$$

In this case $w_{4i\_cal}^{3-4}$, the W3-W4 calibrated longitudinal weight is substituted into equation (12) in order to calculate the Kish effective sample size, and this is then used to calculate the factor detailed in equation (11). After comparing the achieved and effective sample size methods, it transpires that multiplying the calibrated weights by the Kish effective sample size results in weights and therefore estimates with a lower variance, and therefore this method is chosen.

3.3. Properties of the W4 Longitudinal Weights

Table 2 - Properties of longitudinal weights

	***n* (individuals)**	**Mean of weights**	**Coefficient of Variation (CV)**	**Min weight**	**Max weight**
W1-W4	**21,247**	**2,736**	**68.3**	**271**	**13,289**
W1-W3	28,696	2,026	59.6	267	7,000
W3-W4	**33,525**	**1,834**	**72.4**	**87**	**10,093**
W2-W3	31,472	1,870	66.1	155	7,000

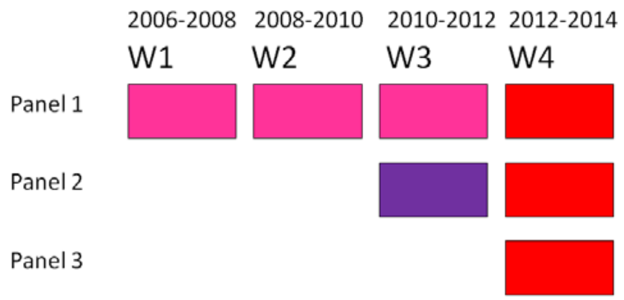
Table 2 examines the properties of the W4 longitudinal weights; the W3 longitudinal weights are also provided for comparison purposes. The number of cases assigned a "survivors" W1-W4 weight is, naturally, smaller than the number of W3 survivors, as further people have dropped out the survey. Therefore, when compared to the W1-W3 longitudinal sample, the W1-W4 longitudinal sample has a larger mean weight as the smaller sample size results in each individual representing a higher proportion of the population.

Inclusion of the W3 top-up panel means the W3-W4 longitudinal sample size is larger than the W2-W3 longitudinal sample size, and its mean weight is slightly smaller. It is interesting to note that the variation in the weights is higher for the W3-W4 longitudinal cases than it is for the W1-W4 survivors, even though the survivors are a smaller subset. As the W3-W4 longitudinal weight consists of two panels, the larger variance is to be expected.

4. Cross-sectional Weighting Strategy

In addition to the two longitudinal weights, a W4 pseudo cross-sectional weight is created. We describe the W4 cross-sectional weight as pseudo because any W4 respondent is assigned a weight regardless of the panel to which they belong (a true W4 cross-sectional weight would only include respondents sampled at the W4 time point). The red boxes in figure 3 combine to give the W4 cross-sectional sample.

Figure 3 W4 cross-sectional multiple panels



The three W4 panels are weighted separately and follow the same general method; each case is assigned a starting weight that is then adjusted for non-response and attrition and then calibrated to known population totals. The three panels are then combined with respect to the effective sample size of each panel in a way similar to that detailed in section 3.2.1. Those who are non-responders for the first time are adjusted for non-response because like attrition, it is a potential source of bias. Cases that leave the survey and then re-enter are adjusted for attrition between non-consecutive waves.

Sections 4.1, 4.2 and 4.3 now consider respectively the cases of those entering the Wave 4 sample initially in Wave 1 (panel 1), Wave 3 (panel 2) and Wave 4 (panel 3).

4.1. Calculating cross-sectional weights for Panel 1

This section concentrates on calculating cross-sectional weights for W4 responding cases that were sampled in Panel 1.

4.1.1 Assigning a starting weight

The first stage in the cross-sectional weighting strategy is to assign a starting weight to all relevant cases.

Previous responders

We first assign a starting weight to all W4 responding cases that have previously responded. A case may have previously responded in W3, W2 or W1. The most recent weight is used as a starting weight, as described below:

- Last responded in W3 -> W3-W4 longitudinal weight
- Last responded in W2 -> W2 cross-sectional weight
- Last responded in W1 -> W1 cross-sectional weight

We now turn our attention to those who have not previously responded.

Joiners to Households

We now assign a weight to those who moved into a W4 responding household since the wave the household last responded. Although the household has previously responded, the individual has not. This type of respondent is assigned a weight using a weight-share method, constructed following Kalton and Brick (1995). This standard approach is based on the W4 household members' starting weights not including the joiners, and sharing these weights between all associated W4 household members.

First, we sum the starting weights of the individuals i, in each household *j,* excluding the W4 joiners. Then we divide this value by the number of individuals in the associated W4 household minus the number of real births, as shown in the formula immediately below, where b_{ij} is the starting weight and N_i^j is the number of individuals in household j.

$$w_{ij} = \sum_{i\varepsilon j=1}^{N_i^j} b_{ij} \Big/ (N_i^j - births_j) \qquad [13]$$

This weight is assigned to all members of a joiner's household. This ensures that all respondents within this type of household have a weight, and that the sum of the weights within that household (and therefore the sum of the weights overall) does not increase when joiners enter the sample (with the exception of actual births, as described below)

Births

Unlike joiners to households, births are a true increase in the population and we want our weights to reflect this. Therefore, births are assigned their "pseudo" mother's weight. That is, births are assigned the weight of the person in the household who is most likely to be the mother. Where there is more than one potential mother in the household, the birth is assigned the average of the potential mothers' weights.

W1 entrants

We now consider cases that respond for the first time in W4 despite being sampled in W1. These are sometimes referred to as "W1 entrants". Entrants' starting weights are their original design weights, d_i constructed as the reciprocal of their selection probabilities.

Table 3 summaries the starting weights assigned to each different type of responder in Panel 1 as discussed above.

Table 3 - Assigning a starting weight for Panel 1 cases

Type of Respondent	Starting weight assigned
W3 and W4	W3-W4 longitudinal weight
W2 and W4	W2 cross-sectional weight
W1 and W4	W1 cross-sectional weight
W4 joiner	Assigned a weight using weight-share method
W4 birth	Assigned pseudo mother's weight
W1 entrant	Design weight

4.1.2 Creating base weights

We are now in a position to join all the different groups that form the W4 sample from Panel 1 and adjust the starting weights; we will refer to an adjusted starting weight as a base weight.

Firstly, the W3-W4 longitudinal responders, births and joiners are joined together and calibrated to the W4 time point population totals; in this case the 2013 mid-year estimates.

Cases that responded in W2 and W4 or W1 and W4, sometimes referred to as "re-entrants", need to be adjusted for attrition. Let us take W2 and W4 cases as an example, where the starting weight is the W2 cross-sectional weight. A number of cases that responded in W2 did not respond in W4, therefore the W2 cross-sectional weight needs to be adjusted for attrition accordingly. This is done using logistic regression following the same method as outlined in section 3.1, where this time W2 information was used to calculate the probability that a case would not respond in W4.

Similarly, cases that responded in W1 and W4 are adjusted for attrition where W1 information was used to inform the logistic regression model.

Thirdly, we adjust the W1 entrants' weights for non-response. Their starting weight is their design weight, which without adjustment assumes that every household selected has responded. This is not the case, therefore we need to multiply the design weights of responding cases by a non-response factor that will compensate for those that did not respond in W4.

Each W1 entrant e is scaled up by the following factor:

$$adj_e = {}^{n_{jz}}/_{m_{jz}} \qquad e \in jz \qquad [14]$$

where n_{jz} is the number of households **sampled** in region j and wealth stratum z and m_{jz} is the number of **responding** households in region j and wealth stratum z, where z = wealthy or not wealthy, using the HMRC definition. For example, if an entrant is from London and is in the stratum of addresses marked as being most wealthy, the adjustment is equal to the number of sampled wealthy households from London divided by the number of wealthy responding households from London.

We now join the W1 entrants and re-entrants to the W3-W4 (already-calibrated) longitudinal responders, births and joiners dataset. The cases that were present in W3 and W4 (and therefore the births and joiners whose weights are based on the W3-W4 weight) were adjusted for non-response as part of the W3-W4 longitudinal weight construction. We therefore need to scale down the weights of this group before joining these cases with the entrants and re-entrants in order not to over compensate for non-response.

The W3-W4 longitudinal responders, births and joiners are scaled down by the following factor:

$$\left(Pop^{w4} - \left(\sum w_i^{e} + \sum w_i^{re}\right)\right) \Big/ \left(\sum w_i^{w3-w4} + \sum w_i^{joiners} + \sum w_i^{births}\right), \quad w_i \in \text{panel 1} \qquad [15]$$

where Pop^{w4} is the W4 time point population total, $\sum w_i^{e}$ is the sum of the entrants' weights, $\sum w_i^{re}$ is the sum of the re-entrants' weights, $\sum w_i^{w3-w4}$ is the sum of the W3-W4 longitudinal responders' weights, $\sum w_i^{joiners}$ is the sum of the joiners' weights and $\sum w_i^{births}$ is the sum of the birth's weights. Table 4 summaries the adjustments assigned to each different type of responder as discussed above.

Table 4 - Adjustments for the starting weight of Panel 1 cases

Type of Respondent	**Adjustment**
W3 and W4, births and joiners	Calibrated and then scaled down
W2 and W4	Attrition adjustment using logistic regression
W1 and W4	Attrition adjustment using logistic regression
W1 entrant	Non-response adjustment using non-response classes

4.1.3 Calibration of base weights

The W4 cross-sectional weights of the W1 panel are then calibrated to the W4 time point population totals; the 2013 mid-year estimates. The aim of the cross-sectional weights is to create a single weight to cover both households and individuals. In order to achieve this, an "integrative calibration" (Lemaitre and Dufour, 1987) approach is used. This results in all people in a household having the same weight, which is also the household weight.

4.2. Calculating cross-sectional weights for Panel 2

This section concentrates on calculating a cross-sectional weight for the W4 **responding cases** that were sampled in Panel 2 (at the W3 time point). The weighting procedure for this panel is the same as the procedure for Panel 1; as Panel 2 cases entered the survey at W3 it is simplified somewhat by the fact that there are no re-entrants from this panel.

4.2.1 Assigning a starting weight

There are fewer types of respondents from Panel 2; W3-W4 longitudinal responders, joiners, births and W3 entrants. The starting weights for these groups are defined in Table 5.

Note that in place of W1 entrants we have W3 entrants. These are cases that were sampled in Panel 2 (the W3 time-point), but respond for the first time in W4.

Table 5 - Assigning a starting weight for Panel 2 cases

Type of Respondent	Starting weight assigned
W3 and W4	W3-W4 longitudinal weight
W4 joiner	Assigned a weight using weight-share method
W4 birth	Assigned pseudo mother's weight
W3 entrant	Design weight

4.2.2 Creating base weights

The starting weights are then adjusted following the same principles outlined in Section 4.1.2. The W3-W4 longitudinal cases, joiners and births are firstly calibrated to the population totals. W3 entrants are adjusted for non-response and then W3-W4 longitudinal cases, joiners and births are reduced by the following factor:

$$\left(Pop^{w4} - \left(\sum w_i^{e} + \sum w_i^{re}\right)\right) \Big/ \left(\sum w_i^{w3-w4} + \sum w_i^{joiners} + \sum w_i^{births}\right), \quad w_i \epsilon\ Panel\ 2 \qquad [16]$$

where the terms used in equation (16) are the same as those defined for equation (15).

4.2.3 Calibration of base weights

The Panel 2 cross-sectional base weights are then calibrated to the 2013 mid-year estimates.

4.3. Calculating cross-sectional weights for Panel 3

As Panel 3 cases were sampled first in W4 of the survey, all W4 responding cases from Panel 3 responded to the survey for the first time in W4. This section will therefore describe how the design weight for the latest panel is calculated, how it is adjusted for non-response and finally calibrated to the population totals.

4.3.1 Calculating the design weight

Panel 3 cases were interviewed for the first time in W4 of the survey; they therefore have no previous weight assigned to them. The starting weight for these cases is therefore the design weight; that is the reciprocal of the selection probability of an address. The two-stage sample design and oversampling of wealthier households is incorporated into the design weight construction as described in section 2.

4.3.2 Adjusting the design weights

Once the Panel 3 cases have their design weights calculated, a non-response adjustment is applied. This is carried out in a similar way to the attrition modelling for the longitudinal weights where a logistic regression model is used to produce the response propensity for each case and is denoted by $\hat{\theta}_{i+}^{r}$. Unlike the attrition models, we have restricted information available to model non-response as no information has been collected from respondents; therefore the only information we can use relates to the address. The variables used to inform the logistic regression model are region, output area classification and wealth indicator. The adjusted design weights d_{i+}^{p3} are calculated as the following:

$$d_{i+}^{p3} = d_{i}^{p3} \cdot \frac{1}{\hat{\theta}_{i+}^{r}} \qquad i \in Panel\ 3 \qquad [17]$$

Finally, these weights are calibrated to the W4 population totals, following the integrative calibration approach, as described in Section 4.1.3.

4.4. Joining the panels for the W4 cross-sectional weight

We are now in a position where we have three separate panels that have been calibrated to the W4 time point. As discussed in section 3.2.1, the panels can be joined with respect to either the achieved or effective sample size of each panel. Both methods were evaluated, where the coefficient of variation of the resulting weights were compared. Table 6 demonstrates how the Kish effective sample size method results in weights with lower variability, making this the preferable method.

Table 6 Achieved versus Kish effective

	Coefficient of Variation of Weights
Achieved	70.3
Kish effective	66.3

Therefore, the final stage of the cross-sectional weighting is to multiply each W4 cross-sectional calibrated weight by the following factor:

$$Kish\ neff = \frac{(\sum_i w_i)^2}{\sum_i (w_i{}^2)} \qquad i \in \text{panel } k \qquad [18]$$

where $k = 1, 2, 3$ and w_i is the calibrated W4 cross-sectional weight.

4.5. Properties of the W4 cross-sectional weights

Table 7 **Properties of cross-sectional weights**

Weight	n	Mean	CV	Min	Max
W4 xs	**46,455**	**1,341**	**66.3**	**36**	**4,847**
W3 xs	49,447	1,207	72.7	69	9,999

Table 7 examines the properties of the W4 cross-sectional weight; the W3 cross-sectional weight is also provided for comparison purposes. The W4 cross-sectional mean weight for W4 is quite similar to that of the W3 cross-sectional weight; the slight increase is mostly caused by the decrease in the sample size from W3 to W4. It also has less variability when compared to the W3 cross-sectional weight, which may be the result of using the Kish Effective sample size method for the first time.

5. Concluding Remarks

WAS has become the prime source of data for estimating wealth across Great Britain.

The weights have two important functions. The first is to account for the unequal selection probabilities used to over-sample people classified as likely to have higher levels of wealth. The second is to adjust for non-response and attrition. Therefore, using the weights helps to reduce bias. However, reducing bias comes at the cost of increasing the variance of the estimates; as the variance of the weights is increased this increases the estimated sampling variance.

Wave 5 of WAS will include the introduction of an additional new panel, the weighting for which can be accommodated within the method principles outlined in this paper.

References

[1] Ashworth, Burnett T. and Smith P. (2012) Weighting the Wave 2 Wealth and Assets Survey. ONS Survey Methodology Bulletin 70, March 2012. Available at: http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-70/index.html [Accessed 10th December 2015].

[2] Kalton G, Brick JM (1995). Weighting schemes for household panel surveys. ONS Survey Methodology Bulletin 21

[3] Lemaître G, Dufour J (1987) ‘An integrated method for weighting persons and families’. ONS Survey Methodology Bulletin 13

[4] Verma V, Betti G, Ghellini G (2007) ‘Cross-sectional and longitudinal weighting in a rotational household panel: applications to EU-SILC’. Statistics in Transition 8

Mailing strategies for optimising response for face to face fieldwork requests

Interviewer led mailings compared with central despatch.

Catherine Grant

TNS

Summary

The Crime Survey for England and Wales (CSEW) is a face to face survey of 35,000 interviews per year. Selected addresses are sent advance notification of the survey in the form of an advance letter.

Letters are sent centrally via second class post one week ahead of the start of fieldwork each month. While this works well for interviewers able to start fieldwork at the beginning of the survey period it may be less effective in cases where the start to fieldwork is delayed due to issues with interviewer availability. An approach used on other studies is to provide the letters to the interviewers instructing them to mail the letters a few days before they attempt face-to-face contact. This approach enables interviewers to stagger the mailing of their letters depending on when they plan to work on particular addresses and may result in a higher recall of the advance information from sampled households and a higher response rate.

Between November 2014 and March 2015 a split sample experiment was conducted on the CSEW to explore the difference in response rates between a central mailing despatch strategy and an interviewer led approach.

The experiment found that the original issue response rate does not differ significantly depending on how the letters are sent. However, the results suggest that allowing interviewers to send out their own letters is on balance more likely to have a positive effect on the original issue response rate than a negative effect. The experiment also indicates that allowing interviewers to send out the letters does not lead to a significant delay in interviewers first attempting to contact respondents. As the despatch method only has a small effect on the response rate it would be beneficial to extend the experiment in order to obtain the statistical power to make the impact estimate more precise.

1. Introduction and background

Advance letters are typical of face-to-face random probability sample surveys conducted in the UK, and these were first introduced to ONS surveys in 1986 and 1987 (Barnes, 1990[1]). These are used to pre-notify sampled households that they have been

[1] Barnes, B. Non-response on government household surveys - A paper prepared for a Workshop on Non-Response at Statistics Sweden, October 1990 and a report of the Workshop, SOCIAL SURVEY DIVISION'S METHODOLOGY PROGRAMME 1990-91

selected for participation and that an interviewer will be calling round to conduct an interview. These letters aim to increase co-operation and (hopefully) participation by 'selling' the study to the residents of each sampled household. These letters usually communicate the purpose of the study, explain what is being asked of the household (or selected individual) and (if appropriate) highlight the incentive that will be offered in exchange for participation. Sending these letters may also be beneficial to a study by increasing the self-confidence of interviewers, providing greater legitimacy to their work and by allowing them to refer to the letter when making contact (Groves, 2004)[2]. There is evidence that advance letters can improve co-operation rates, albeit generally marginally (e.g. Lynn, Smith and Turner, 1998[3], and De Leeuw et al., 2007[4])[5].

Advance letters are used as standard practice on the CSEW. Interviewer assignments are issued on a monthly basis and the letters are despatched centrally via second class post one week before fieldwork begins. This approach is used for a number of random probability sample face-to-face surveys carried out in the UK.

An alternative approach is used on other studies such as the European Social Survey, the DCMS Taking Part survey and the Cabinet Office's Community Life survey. With this approach, interviewers are sent all the letters for their assignment in a single pack and are instructed to post them out a few days before they attempt face-to-face contact. The hypothesis is that interviewers can ensure that their first visit is always shortly after the mailing has been received - when potential respondents are most likely to remember the letter and its contents – and will obtain higher cooperation rates as a result. This is a particularly useful for interviewers who are unable to start working immediately. Interviewers can also send out their letters in batches depending on when they plan to first make contact at each address in their assignment.

However, an alternative hypothesis is that, by giving interviewers autonomy over when letters are sent we encourage some to delay starting on their work. This may lead to over-compressed fieldwork, lower response rates and a greater degree of inefficient 'reissuing' of initially unproductive cases. In contrast, a centralised despatch may prompt interviewers to begin working on their assignments as soon as possible so that they get the maximum benefit from the advance mailings.

It is important to note that the method of administering mailings also has other cost implications, which need to be considered when deciding on the best approach to use. Allowing interviewers to send out the letters entails higher costs: (i) additional administration time to batch up the letters into the correct packs, (ii) additional

---

www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-28/survey-methodology-bulletin-28.pdf

2 Groves, Robert M., Survey Errors and Survey Costs, New Jersey, John Wiley & Sons, Ltd., 2004

3 Lynn, Peter; Smith, Patten; Turner, Rachel Assessing the effects of an advance letter for a personal interview survey. Journal of the Market Research Society 40.3 (Jul 1998): 265-272

4 De Leeuw, E. , Callegaro, M., Hox, J., Korendijk, E. & Lensvelt-Mulders, G. The Influence Of Advance Letters On Response In Telephone Surveys - A Meta-Analysis. Public Opinion Quarterly, Vol. 71, No. 3, Fall 2007, pp. 413–443

5 It is also worth noting that some literature suggests that advance letters can in some circumstances have a negative effect, for instance, by increasing refusal rates (Groves, 2004) there is little actual experimental evidence for this.

postage costs (due to additional weight of the interviewer packs) and (iii) no volume-based discounted postage rate (as currently obtained for the central despatch method). Consequently, this method is not cost-free and should be evaluated for its benefits before introducing it to the CSEW.

So far as we know there are no studies published which determine which of these two approaches is the most effective in maximising the original issue response rate in face-to-face surveys. In order to investigate this, an experiment was conducted with the CSEW.

2. Design

A randomly selected half of all assignments was subject to central letter despatch while for the other half the letters were included in the interviewer packs for interviewers to mail themselves. Interviewers were informed of the experiment and how it should be administered.

The experiment was conducted on the set of CSEW addresses that was issued between November 2014 and March 2015 inclusive. The experiment was designed to have sufficient power (80%) to detect an effect of 2.5 percentage points; an effect deemed to be large enough to be notable and worth changing the survey protocol for. A period of five months was calculated as being sufficient to detect this change (22,500 issued addresses, 11,250 per arm). This meant that with an assumed design effect of c.2 a difference of more than 2.4 percentage points (from a base percentage of 62% - a typical original issue response rate) would lead to a positive significance test result ($p<.05$) with 80% probability.

In practice, the power of the experiment was less than expected and the two reasons for this are outlined below.

Firstly, the design effect was larger than originally expected; prior to the study it was estimated that the design effect would be c.2, but the final design effect for the overall response rate calculation was 2.96 for centrally despatched addresses and 2.6 for interviewer despatched addresses. The clustering of the sample was the main component of the design effect, and we believe that the higher than expected level of intra-cluster homogeneity is due to there being substantial interviewer effects as well as area effects.

Secondly, it is important to note that from the beginning of 2015 – and for several months afterwards - the fieldwork for the CSEW was carried out partly by GfK, temporarily supporting the primary contractor TNS BMRB. However, the experiment was carried out solely among those assignments administered by TNS BMRB and this reduced the size of the experiment and the power to detect a difference between the two methods.

In total, 562 assignments were issued as part of this experiment (comprising 18,631 addresses in total). The allocation of these assignments (and their addresses) to the different experimental cells is shown below.

Table 1 Design of the experiment

Experimental cell	Assignments		Addresses	
	N	%	N	%
Central despatch	285	50.7%	9,487	50.9%
Interviewer despatch	277	49.3%	9,144	49.1%

Each arm of the experiment ended up having an effective sample size of c.3,000. The final experiment only had the power to detect a difference of 2.5 percentage points (from the estimated baseline) 51% of the time (rather than 80% as originally intended) and instead had 80% power to detect a 3.5 percentage point difference from the baseline.

3. Objectives

The specific objectives for this experiment were to determine whether:

1. The original issue response rates differ depending on the advance mailing strategy.
2. To examine when the first call was made to each address in order to ascertain whether allowing interviewers to post out their own letters leads to a delay in the beginning of fieldwork

Prior to the experiment it was hypothesised that results may vary according to:

- **Region** - the response rate for CSEW varies geographically and it was hypothesised that the manner in which letters are despatched may have a larger impact in some areas than in others.
- **Interviewer tenure** - providing interviewers with greater autonomy over their workload (i.e. by giving them flexibility over when their letters are sent out) may have a different effect depending on the experience level of each interviewer. For instance, those that had less experience may find it more difficult to manage their workload and to send out letters efficiently. For this analysis, the interviewers who worked on the experiment were classified into quartiles based on the length of time they had worked as an interviewer for TNS[6].

The results have been broken out for these two groups in the following analyses.

4. Analysis

4.1. Original issue response rate

Analysis focuses on the response rate to the original issue fieldwork period rather than the final response rate; there are two main reasons for this:

[6] This is not a perfect measure as some interviewers who are new to the TNS panel may in fact have many years of experience working as an interviewer in another field force.

1. Advance letters sent to potential respondents are only likely to have a short term effect which will be visible in the eight week original fieldwork period but not during the re-issue period
2. After the original fieldwork period, non-contacts and soft refusals are reissued back into field and this process is likely to smooth out any differences caused by the different treatment of the advance letters

The random allocation of assignments to the experimental cells means a difference can be identified just by comparing the mean response rates. The analysis took into account the complex sample design of the CSEW: the stratification by police force area, the geographic clustering and the variations in address sampling probability.

The response rate for the two experimental groups is shown in the following table – the serials where the letter was despatched by an interviewer had an original issue response rate of 63.3% and those with a central despatch had an original issue response rate of 61.2%. The difference is 2.1 percentage points but the standard error of the difference between the two proportions is 1.3 percentage points, giving a 95% Confidence Interval of -0.4%pts to +4.6%pts. This means that the difference of +2.1 percentage points was not quite found to be significant at the 95% level (T score of 1.65 and a p value of 0.10 assuming a two-tailed test). Nevertheless, the null hypothesis of no effect is not strongly supported by this data. It would be reasonable to expect the interviewer despatch method to generally lead to higher response rates than the central despatch method even if the statistical support for this expectation is modest.

Table 2 Original issue response rate (excluding deadwood) for both experimental cells

	Mail out method		Interviewer despatch minus Central despatch
	Central despatch	Interviewer despatch	
Estimate	61.2%	63.3%	+2.1%
Lower 95% Confidence Interval	59.3%	61.5%	-0.4%
Upper 95% Confidence Interval	63.0%	65.1%	+4.6%
Standard error	0.9%	0.9%	1.3%
Base (assignments)	285	277	
Base (issued addresses minus deadwood)	8,666	8,437	

The results are broken down by region in the table 3. The original issue response rate was higher in seven of the nine regions when interviewer mailings were used instead of central despatch. None of the differences were found to be significant at the 95% level (though Wales is very close) but there is some *consistency* in the direction of the difference. In reality, this experiment lacks the power to detect regional variation when the effect is modest. The noise of random sampling error is too dominant to detect the signal of a meaningful effect.

Table 3 – Original issue response rate for both experimental cells by region

	Central despatch	Interviewer despatch	Difference	Standard error of the difference	T Score	P value
North East	66.9%	71.6%	+4.7%	5.7%	0.83	0.41
North West	62.0%	65.8%	+3.8%	4.2%	0.90	0.37
Yorkshire & Humberside	65.9%	62.5%	-3.4%	3.8%	0.89	0.37
East Midlands	63.5%	66.7%	+3.2%	3.3%	0.97	0.33
West Midlands	63.5%	64.6%	+1.1%	5.7%	0.19	0.85
East of England	63.7%	59.8%	-3.9%	3.1%	1.25	0.21
London	45.8%	49.8%	+4.0%	4.1%	0.96	0.34
South East	62.5%	64.2%	+1.7%	3.2%	0.53	0.59
South West	61.1%	65.4%	+4.3%	3.1%	1.38	0.17
Wales	62.7%	71.8%	+9.1%	4.7%	1.93	0.05

Table 4 shows the results broken down by interviewer tenure; this bivariate analysis fits with the prior hypothesis and suggests that there may be a significant impact for interviewers with a long tenure (a higher response rate where they send out the letter), whereas no effect is observed for those with a shorter tenure. However, when we use a regression model to test whether the impact of the experiment is mediated by interviewer tenure (table 5) we find that the interaction term is not significant (p=.24) and therefore we cannot be confident that the impact varies with interviewer tenure.

Table 4 Original issue response rate for both experimental cells by interviewer tenure

	Central despatch	Interviewer despatch	Difference	Standard error of the difference	T Score	P value
Quartiles 1-3 - More experienced interviewers	60.1%	63.2%	+3.1%	1.5%	2.03	0.04*
Quartile 4 - Least experienced interviewers	64.3%	63.7%	-0.6%	2.6%	0.24	0.81

Table 5 Regression to test impact of experiment by interviewer tenure

	Tests of model effects			
Source	**df1**	**df2**	**Wald F**	**Sig.**
(Corrected Model)	3.000	518.000	2.002	.113
(Intercept)	1.000	520.000	270.858	.000
Interviewer tenure[7]	1.000	520.000	2.238	.135
Mail out method[8]	1.000	520.000	0.591	.442
Interviewer tenure * Mail out method	1.000	520.000	1.388	.239

Dependent Variable: Original issue response rate (reference category = .00)
Model: (Intercept), Interviewer tenure, Mail out method, Interviewer tenure * Mail out method

4.2. First contact attempt made to each address

The interviewers working on the CSEW use an Electronic Contact Sheet (ECS) to log all of the contact attempts which they make at each address. The ECS records the time and date of each visit as well as other information such as the outcome. This paradata has been used to calculate the number of days after the official start of the fieldwork period that interviewers first attempted to make contact at each address[9].

The median number of days for the first contact was found to be 13 days, and this did not vary between the two mail-out methodologies. There is a difference of 1.16 days between the mean for each approach, with addresses sent a mailing from a central location having a lower mean of 16.44 days, however this was not found to be significantly different (T score of 1.18, p value 0.24) from the mean for interviewer despatch (17.60).

Table 6 shows – for both experimental cells - the average number of days before the first contact attempt; as for the earlier analysis the standard errors calculated for the mean take into account the clustering, stratification and variation in sampling probabilities.

A linear regression model was also used to determine whether the impact on timing of first contact varied according to region or interviewer tenure. The model used the number of days after the beginning of fieldwork that contact was first attempted as the dependent variable and the mail out method, region and interviewer tenure as predictors. As shown in table 6, the interaction terms for region and interviewer tenure when crossed with the experiment condition were not found to be significant – demonstrating that there was no variation in the mail out method effect between regions or between interviewer tenures.

[7] Included as a factor with two categories: “Quartiles 1-3 - More experienced interviewers” and “Quartile 4 - Least experienced interviewers”

[8] Included as a factor with two categories: “Interviewer despatch” and “Central despatch”

[9] There was some missing data – with about 100 addresses issued over the course of the experiment missing call record data. These have been omitted from the analysis.

Table 6 Average number of days after official fieldwork start that contact was attempted at addresses

	Mail out method		Difference
	Central despatch	Interviewer despatch	(interviewer despatch minus central despatch)
25th Percentile	6	7	-1
Median	13	13	0
75th Percentile	23	25	-2
Mean	16.44	17.60	+1.16
Lower 95% Confidence Interval of Mean	15.11	16.20	+1.09
Upper 95% Confidence Interval of Mean	17.78	18.99	+1.21
Standard error of Mean	0.679	0.709	0.982
Base	9,434	9,095	

Table 7 Regression to test whether effect of mail out approach varies by region or interviewer tenure

Tests of model effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	25.000	496.000	2.212	.001
(Intercept)	1.000	520.000	1251.759	.000
Mail out method[10]	1.000	520.000	.824	.364
Interviewer tenure[11]	3.000	518.000	3.869	.009
Region[12]	9.000	512.000	4.560	.000
Mail out method * Interviewer tenure	3.000	518.000	.257	.856
Mail out method * Region	9.000	512.000	.343	.960

a. Model: Number of days contact first attempted = (Intercept) + Mail out method + Interviewer tenure + region + Mail out method * Interviewer tenure + Mail out method * region

5. Conclusion

The results suggest that allowing interviewers to send out their own letters is on balance more likely to have a positive effect on the original issue response rate than a negative effect when compared to the central despatch method. However, given the findings were not statistically significant, it is still possible that the effect is zero or too small to warrant the additional costs. The findings do however show that allowing interviewers to send out their own mailings does not lead to a significant delay in interviewers first attempting to contact respondents. This suggests that there is little risk that the original response rate would decline should the despatch method be changed.

[10] Included as a factor with two categories: "Interviewer despatch" and "Central despatch"

[11] Included as a factor with four categories: "Quartile 1 - Most experienced interviewers", "Quartile 2", "Quartile 3" and "Quartile 4 - Least experienced interviewers"

[12] Included as a factor with ten categories, one for each former Government Office Regions in England and one for Wales

The primary consideration for CSEW therefore relates to cost and the scale of survey means that this change would lead to a significant increase in the cost of processing the advance letters. It is therefore important to ensure that there is definitely a beneficial effect before implementing this change. The final approach to be used for CSEW is still being considered based on this evidence, and ONS is considering extending the experiment to obtain more statistical power to detect a smaller effect. For instance, to detect an effect of 2.5 percentage points the experiment would need to be extended by another c.17,000 issued addresses (i.e. about four months of the CSEW).

These learnings may also apply to other surveys, although it is important to caveat that the impact of allowing interviewers to send out letters may vary depending on such factors as the envelope used, the survey sponsor and the topic of the survey. Given that there is no strong evidence that interviewer posting has a positive impact on the response rate, other metrics such as cost and quality considerations should be used to decide on the mailing protocol to be used for any given survey.

Future research could be used to build upon these findings; in particular it would be valuable to record the date on which interviewers post out their letters in order to expand the analysis of outcomes that can be performed with respect to advance mail out strategies.

Labour Force Survey (LFS) follow-up surveys: Examples and methodological considerations

David Ainslie, Matt Greenaway, Gareth Rusgys and Tim Vizard.

Abstract

This paper highlights how the Office for National Statistics (ONS) has identified and delivered an effective model of data collection; using respondents who did not object to re-contact at their final interview of the Labour Force Survey (LFS) as a sample frame for future follow-up surveys.

The paper highlights two examples where this approach has enabled ONS to target populations of interest to quickly and efficiently provide reliable data that has helped inform public policy decisions:

- the European Health Interview Survey (EHIS), conducted on behalf of the Department of Health (DOH) and required under European legislation and
- the Survey of the Self-Employed (SES), conducted on behalf of the Department for Business Innovation and Skills (BIS)

The paper then summarises the opportunities and challenges of conducting such surveys, discussing the sampling and weighting methodology that can be used to control a key challenge - the potential for non-response bias.

1. Introduction

The Labour Force Survey (LFS) is the largest household survey conducted in the UK, with around 40,000 responding (or imputed) households per quarter[1]. The LFS utilises a rotating panel design, meaning that most households, once sampled, are interviewed for five consecutive quarters, known as 'waves'.

A question is included in the LFS such that at their final LFS interview ('final wave'), respondents are asked if they would not object to re-contact from ONS to take part in future surveys. Respondents who do not object provide a potential sample frame for follow-up of approximately 9,000 households or 20,000 individuals per quarter[2].

[1] LFS User Guidance Volume 1 Guidance and Methodology: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance

[2] In addition to respondents who have completed five LFS waves and are asked if they would not object to re-contact at their final wave, the potential sample frame for follow up includes households containing only economically inactive people aged 75 and over (who are only…

ONS has utilised this follow-up sample frame for internal purposes, such as to conduct the LFS Dress Rehearsal or to provide a sample for cognitively testing survey questions.

Use of this LFS follow-up sample frame has also enabled ONS to conduct follow-up surveys that efficiently target populations of interest and meet tight deadlines to deliver quality and timely data that assist UK government departments and Eurostat (the statistical office of the European Union) in filling gaps in the evidence base used to inform policy. This paper highlights two examples of this approach, and summarises the opportunities and challenges posed by this kind of follow-up survey.

Two recent examples of ONS utilising the LFS follow-up sample frame in such a manner are:

- the European Health Interview Survey
- the Self-Employed Survey

The European Health Interview Survey (EHIS)

The EHIS is required to be conducted by member states under European legislation and was conducted by ONS on behalf of the Department of Health (DoH) in England, and the devolved health bodies in Wales, Scotland and Northern Ireland.

The purpose of EHIS is to gather information on the health status, health determinants and health services used by households across the UK. The UK Government and European Union use the results of each wave of EHIS to monitor and inform health policies including strengthening the European communicable diseases alert system.

The Self-employed Survey (SES)

The SES was commissioned as a one-off survey by the Department for Business Innovation and Skills (BIS) as part of their Understanding Self-Employment programme.

The purpose of the SES was to gather more-detailed information on the diverse circumstances, motivation and challenges facing those in self-employment in the UK. The UK Government used the SES to inform self-employment policy and add to an independent review of self-employment conducted for UK government by Julie Deane OBE in 2015 to 2016.

A summary of the key features of the EHIS and the SES is provided in Table 1.

…required to complete one wave of the LFS), and respondents to the Local Labour Force Survey (LLFS) boost. Respondents in both of these groups are also asked if they would not object to re-contact. For more information on these groups see: LFS User Guidance Volume 1 Guidance and Methodology, pages 18 to 20:

https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance

Table 1: EHIS and SES summary

Survey	**EHIS, 2013-2014, Great Britain**[3]	**SES, 2015, Great Britain**
Sample Design	Sample of 22,818 *households* in Great Britain where • 1 or more household members had not objected to re-contact by ONS to take part in future surveys when completing their final wave of the LFS. Household members who had previously refused to take part, were non-contactable or who had refused permission to re-contact on the LFS previously were invited to take part in the survey only if consent was volunteered. Any new adult household members were invited to take part i.e., anyone who had moved into the household, or become 16, since the last LFS interview.	Sample of 2,503 *individuals* who had • identified as self-employed in their main job • not objected to re-contact by ONS to take part in future surveys or • not answered the re-contact question and someone in their household had not objected to re-contact, when completing their final wave of the LFS.
Questionnaire Design	Model questionnaire covering health element of the survey provided by Eurostat. Supplemented with a series of standard questions used by ONS to gather the data required to derive core social variables required for delivery to Eurostat. Complete interview expected to last average of 45 minutes. 80% Computer Assisted Telephone Interview (CATI) and 20% Computer Assisted Personal Interview (CAPI) formats.	Developed by ONS Social Survey Division in consultation with BIS from an initial set of questions. 50 mostly closed questions relating to aspects of self-employment. Complete interview expected to last an average of 20 minutes. Computer assisted telephone interview (CATI) format only.
Pilot	1 week pilot in January 2013 to test the questionnaire, advance letters, interviewers' briefing materials, and systems used to gather, extract and process collected data.	None.
Fieldwork	April 2013 to March 2014.	September 2015 to October 2015.
Response Rate	Overall response rate of 62%	Overall response rate of 57%
Dissemination	Summary of results and reference tables: Health indicators for the United Kingdom, ONS, 2015. Dataset and Quality report provided to Eurostat. Dataset supplied to UK Data Service in January 2016. To encourage the widest use of the dataset a number of treatments were applied to the version deposited to permit it release as an End-User Licence dataset.	Summary of results and survey technical report: Understanding Self Employment, BIS, 2016. Dataset to be supplied to the UK Data Service in 2016.

[3] Note that in Northern Ireland, EHIS, 2013-2014, was conducted using a different methodology

2. Opportunities and Challenges of using the LFS follow-up sample frame

The LFS follow-up sample frame provides a valuable resource to deliver social survey requirements. Some of the opportunities and challenges associated with the use of this sample frame relate to:

Detailed information available on the sample frame

A key practical constraint when designing and undertaking social surveys is the availability of a good-quality sampling frame.

Most ONS UK household surveys use the Postcode Address File (PAF), a list of addresses maintained by the Royal Mail, as a sampling frame. The PAF offers good coverage of the UK private household population and provides a robust basis for most ONS social statistics. However, using the PAF poses a number of challenges for smaller, one-off surveys.

Using the PAF as a sampling frame for surveys where the target population is a particular subset of the private household population is particularly challenging as the PAF contains very little information about the occupants of an address. This often means an expensive 'screener' survey is required since subsets cannot be directly identified using this sampling frame.

Even for surveys where the target population is most of the private household population, the PAF suffers from some over-coverage, mostly vacant households or businesses. Typically, around 90% of sampled addresses will be used as a primary residential address.

Phone numbers are not available for the majority of addresses on the PAF, meaning more expensive face-to-face surveys are often required. The Labour Force Survey, for example, is mostly face to face at wave 1 and mostly telephone at later waves using telephone numbers collected during the wave 1 interview.

By using respondents who have consented to being re-contacted on the LFS as a sampling frame, it is possible to avoid many of these issues. Information collected on the LFS for these respondents, results in a lot of information we can utilise in the design and conduct of the survey. In particular, we can effectively identify sub-groups of interest and conduct primarily telephone surveys instead of relying on the more expensive face-to-face mode.

For example, the target population for the SES was self-employed individuals only. There is no way to identify this subgroup on the PAF, meaning that a PAF-based survey would need to have employed a sample several times larger and a 'screener' survey to remove non-self-employed individuals from the sample. The survey would also probably have had to be primarily face to face instead of entirely telephone, which would have been considerably more expensive.

Response Rates

Long-term trends in household survey response rates for social surveys are downwards as a result of the increasing refusal to participate and household non-contact. The LFS follow-up survey allows ONS to contact adults who have consented to re-contact in the future, and therefore are more likely to take part in a follow-up survey. ONS experience with the European Health Interview Survey found that response was 62%, a relatively high rate for a predominantly telephone interview. This helps to reduce fieldwork costs.

Interview Length

The LFS follow-up sample frame allows the ability to match data previously collected on the LFS to new survey collection data. In a time when all statistical providers are seeking to make use of existing administrative data sources to maximise data quality and reduce respondent burden, the LFS follow-up design enables ONS to make use of existing data available.

Using EHIS as an example, ONS were able to use information previously collected from respondents on core LFS questions. This allowed for complex questions on employment, educational attainment and other core demographic characteristics of the household to be checked with participants only, rather than re-asked. This resulted in a reduction in interview time and respondent burden.

The use of previously collected LFS data was achieved using a method proven on other ONS longitudinal surveys. ONS took the opportunity to design and programme a Computer-Assisted Personal Interview (CAPI) and Computer-Assisted Telephone Interview (CATI) using Blaise questionnaire software, which has the functionality to pre-load and securely store data into the questionnaire, and populate questions with data obtained from previous surveys.

Non-response bias, sample design and weighting

The key challenge of using the LFS follow-up sample frame is that the sample is to an extent 'self-selecting', in that it only allows the approach of those who have both responded to the LFS and have then not objected to re-contact. Whilst this introduces a potential source of non-response bias, ONS is able to utilise a number of statistical methods aimed at controlling non-response bias using weighting. The remainder of the paper outlines methods for controlling non-response bias, how these have been implemented by ONS to date on the SES and EHIS, before considering methods that could be used in the future.

3. Non-response bias

Sources of non-response bias on LFS follow-up surveys

LFS follow-up surveys include a number of stages of non-response:

- non-response to the LFS. This can be either non-response at first interview ('wave 1') or individuals subsequently dropping out at waves 2-5, known as 'attrition'. As of quarter 3 2015, wave 5 LFS

response rates are around 41%, attributable to a combination of wave 1 non-response and attrition[4]
- non-consent to follow-up when interviewed at final wave of the LFS
- non-response to the follow-up survey itself

This larger than normal scope for non-response means non-response bias, stemming from responders being systematically different from non-responders at any of these stages, is a particular methodological concern.

Statistical methods for controlling non-response bias

Statistical methods for controlling non-response bias in social surveys using weighting are well developed (see for example, Särdnal, Swensson and Wretman, 1992, Chapter 15). There are two general approaches, which may be used individually or together; sample-based weighting and population-based weighting.

Sample-based weighting involves using sample data about responders and non-responders to estimate response propensities. These response propensities can then be used when calculating survey weights to ensure the weighted respondents are representative of sampled individuals. A simple example might be that if we can see from our sample and response data that the owner-occupiers we sampled were more likely to respond than non-owner-occupiers, we can utilise a weighting method which gives the relatively fewer responding non-owner-occupiers a larger weight.

Sample-based weighting will reduce non-response bias if the variables used to carry out the adjustment are strong predictors of response and are correlated with survey outcome variables. There is, however, a cost, as it may increase the variance in the weights and therefore inflate standard errors (Elliot, 1991).

Sample-based weighting is usually constrained by the fact that, as previously outlined, most ONS social surveys utilise the PAF as a sampling frame, and there is relatively little useful information on the PAF about non-responders. For example, we do not know the tenure status of LFS non-responders at wave 1, and so couldn't use tenure to estimate response propensities. However, this is not always the case for follow-up surveys. We have LFS data for all final wave LFS responders, and wave 1 LFS data for all individuals who drop out of the LFS or do not consent to follow-up. This means that there is considerably more scope for carrying out sample-based non-response adjustments.

Population-based weighting involves using population size data when calculating survey weights. For ONS social surveys, this stage is typically carried out using population projections, which use census and administrative data to provide accurate estimates of the UK population size broken down by age, sex, and low levels of geography. This stage will remove non-response bias related to age, sex or geography under similar conditions to sample-based weighting, and additionally may reduce

[4] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

standard errors. A simple example might be that we can ensure the age distribution in our responding sample matches the age distribution in the overall population.

The variables used in population-based weighting are normally limited to age, sex and geography, as these are usually the only variables for which good quality population estimates are available and which are collected on most social surveys. However, if good quality survey estimates are available, these may also be used in population-based weighting. This may add value if a larger-scale survey is run that collects some of the same variables as the smaller-scale survey, estimates from the larger survey may be used in the population-based weighting of the smaller survey. This is usually the case with LFS follow-up surveys, since the LFS is a very large survey, and LFS variables can be matched onto follow-up survey datasets.

In summary, while there is more scope for non-response bias with follow-up surveys, there are also more powerful methods for controlling non-response bias, as we have information about both responders and non-responders which can be used in both sample-based and population-based weighting.

Weighting methods used for EHIS and SES

Both the EHIS and the SES used *two-phase population-based weighting,* in addition to standard steps to account for selection probabilities.

The EHIS weighting utilised population estimates for:

- population size by age (five-year age bands) by sex for all GB – from ONS population projections
- population size by age (ten-year age bands) by sex for England, Scotland and Wales – from ONS population projections
- population size by region – from ONS population projections
- population size by economic activity status (employed/unemployed/inactive) – estimated from the Annual Population Survey (APS)[5]

Investigations took place into using sample-based weighting, but this had a small impact on estimates and increased the standard errors of estimators, and we therefore did not use sample-based weighting. Practical difficulties with matching datasets also limited the scope of sample-based adjustments we could use, an issue which will be fixed for future follow-up surveys.

The SES utilised population estimates for:

- self-employed population size by region – estimated from APS
- self-employed population size by age (fifteen-year age-bands) – estimated from APS

[5] For more information on the Annual Population Survey (APS) please see LFS user guidance Volume 6 APS User Guide:
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance

- self-employed population size by sex and over/under 55 – estimated from APS

It is difficult to evaluate the efficacy of weighting in removing bias, as in most cases we do not know the ‘true’ unbiased estimate. We can, however, compare to other survey estimates, in particular from surveys with fewer possible ‘layers’ of non-response and therefore less scope for non-response bias. The table below contains EHIS un-weighted and weighted estimates compared with equivalent APS estimates for self-reported health. Note that the EHIS estimates here were calculated using an early version of the dataset and so may not be identical to published estimates.

Table 2: EHIS and APS estimates (April 2013 to March 2014)

		EHIS Estimates		APS Estimate
Country	**Self-reported health**	**Unweighted**	**Weighted**	**Weighted**
Eng	1 (very good)	31.0%	35.8%	39.0%
Eng	2	40.4%	40.0%	38.4%
Eng	3	20.5%	17.4%	16.1%
Eng	4	6.2%	5.3%	4.9%
Eng	5 (very bad)	1.8%	1.6%	1.5%

In all cases the EHIS weights move the estimates in the direction of the APS estimates. The age element of the weighting has the most impact, younger individuals are less likely to respond, and tend to have better health. However, other elements to the weighting also have an impact. In particular, if the ‘two-phase’ element of the weighting was not used, that is, if age, sex, and region were used without economic activity, then the EHIS estimates would be more different to the APS estimates, with a drop in the estimated proportion of individuals with ‘very good’ health. This suggests that the weighting method is removing at least some non-response bias from estimates.

Weighting methods for future follow-up surveys

While there is more scope for non-response bias with follow-up surveys, there are also powerful methods for controlling non-response bias. ‘Two-phase’ population-based weighting appears to have worked well in recent surveys, and has the added advantage of being relatively straightforward to implement, which is important as follow-up surveys typically have a quick turn-around. The variables to be utilised will vary according to context, and should utilise recent results on LFS attrition (see Lacey, Cooper and Greenaway, 2016).

Sample-based weighting appears to make relatively little difference ‘over and above’ population-based weighting in some scenarios. However, this will vary according to context, and improvements in dataset-matching may allow more powerful sample-based adjustments to be used in future.

4. Conclusion

We have summarised the opportunity to use the detailed information available on the LFS follow-up survey sample frame to target sub-populations, improve response rates, reduce interview length and fieldwork costs. The paper has highlighted two recent surveys where this approach has been successfully used to deliver robust, cost-effective statistics. ONS are planning to use this approach to deliver a number of future surveys.

It is important to account, as far as possible, for the additional scope for non-response bias in follow-up surveys. Fortunately, the LFS provides a rich source of information for most non-respondents, allowing for robust estimation methods that account for non-response. Further work will continue to develop these estimation methods.

References

Elliot, D. (1991) Weighting for non-response, OPCS

Lacey, A., Cooper, D., and Greenaway, M. (2016) "Investigating Attrition on the Labour Force Survey" Survey Methodology Bulletin 72, ONS

Särndal, C.E., Swensson, B., Wretman, J. (1992) Model Assisted Survey Sampling, Springer

Methodology Advisory Service (MAS)

The Methodology Advisory Service is a service of the Office for National Statistics (ONS); it aims to spread best practice and improve quality across official statistics through methodological work and training activity. The ONS has about one hundred methodologists - highly qualified statisticians and researchers; their primary role is to provide expert support, advice and methodological leadership to the ONS in producing and analysing National Statistics.

Methodology staff are arranged into Centres of Expertise, each comprising a team of specialists who keep abreast of research and developments in their area of expertise through contacts with academia, other national statistical institutes and the wider research community. Many of these Centres have international reputations and present research and applied work at conferences and at other meetings of experts in their fields. Examples of these centres are Sample Design and Estimation and Time Series Analysis.

The Methodology Advisory Service has a remit to extend the services of ONS methodologists beyond ONS into other public sector organisations. Every year, MAS carries out projects with customers addressing a wide range of statistical requirements. As well as calling on methodology staff, MAS can also draw on the wider expertise of statisticians, researchers and subject area specialists across the ONS. Further expertise is available though links with Universities.

Contact ***MAS@ons.gov.uk***

GSS Methodology Series

Latest reports in the GSS Methodology Series:

37. *Estimating alcohol consumption from survey data: updated method of converting volumes to units*
38. *100 Years of the Census of Production in the UK, Paul Smith*
39. *Quality of the 2010 Electoral Register in England & Wales, Neil Hopper*
40. *Modelling sample data from smart-type electricity meters to assess potential within Official Statistics, Susan Williams and Karen Gask*
41. *Using geolocated Twitter traces to infer residence and mobility, Nigel Swier, Bence Komarniczky and Ben Clapperton*

Reports are available from:

http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

Forthcoming Courses

1. GSS Statistical Training Programme

A series of government specific short courses (between 0.5 and 2 days in length) delivered by methodological experts in the field. These courses are delivered at ONS sites in London, Newport and Titchfield.

For further information on the available courses see the Statistical Training Service prospectus:

https://gss.civilservice.gov.uk/wp-content/uploads/2013/02/GSS-Training-Prospectus.pdf

or contact ***statistical.training.enquiries@ons.gov.uk***

Latest timetable:

https://gss.civilservice.gov.uk/wp-content/uploads/2015/10/Statistical-Short-Course-Programme-v6.pdf

2. MSc in Official Statistics

This MSc programme has been developed jointly by the University of Southampton with National Statistics to cover the core skills and knowledge needed by professional government statisticians working in the UK and in other countries:

More information: www.southampton.ac.uk/moffstat

Enquiries

The Survey Methodology Bulletin is usually published twice a year, in Spring and Autumn. Copies of many previous editions are available electronically at:

http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/index.html

If you would like to be added to the distribution list please email ONS Methodology at:

methodology@ons.gsi.gov.uk

Or write to us at:

Survey Methodology Bulletin
Room 1.156
Office for National Statistics
Cardiff Road
Newport
NP10 8XG

ons.gov.uk
visual.ons.gov.uk