Summary of ONS project to look at DWP Benefit Fraud and Error estimators and confidence Intervals

Background

DWP Fraud and Error Measurement and Analysis (FEMA) team are responsible for measuring fraud and error in the DWP benefit system. The DWP benefit expenditure budget is around £160 billion, of which around £3 billion is estimated to be overpaid. We produce a bi-annual National Statistics publication to report estimates of fraud and error. These estimates are based on a sample survey which is run on a continuous basis by an in-house data collection team. Within our publication we provide a central estimate of fraud and error for each benefit and associated confidence intervals calculated at a 95% level of confidence. Our central estimates are subject to high levels of variance and therefore our confidence intervals (CIs) are large. This is particularly problematic when comparing year on year changes.

We are investigating methods to improve our CIs because the results are used by the Department to measure performance against specified fraud and error targets. In addition, we are also undergoing a challenging resource reduction agenda across the department. A narrowing of our confidence intervals may allow us to reduce our sample size and hence operational resource.

We applied for a Quality Improvement Fund (QIF) as we welcomed advice, guidance and support from the Office of National Statistics (ONS). Our QIF bid was accepted and we asked ONS to look at the different methods that would be appropriate to use to improve how we calculate our confidence intervals. This has been highlighted as a suggested area of improvement by the UKSA assessment team. An improved estimator and confidence intervals would hopefully reduce the variance associated with our estimates and thus enable our users to more accurately compare change over time. This could also enable us to look at reducing the size of the sample (if we have narrower confidence intervals) and hence reduce the DWP resource that is being spent on collecting the sample data.

The reasons for ONS advice

There were various reasons for asking ONS for help and these are discussed in this section. One point to note is that the actual section of ONS that we worked with is known as the Methodology Advisory Service (MAS). The key reasons for applying for a QIF grant were because we wanted:

- An independent check of our current approach
- A comparison of our methods with other departments
- A comparison of formulae-based variances and simulation based variances.
- To identify the best approach for future measurement, particularly Universal Credit.

Each of the above points is discussed below.

Independent check of current approach

The estimator that grosses-up the sample data to provide national estimates was chosen many years ago. And although that may have been appropriate at the time, we wanted to get some independent advice on the best option for the current datasets. This is because there are various estimators that could have been used to gross-up the sample so we wanted MAS to investigate the options. The options that MAS considered were:

- Horvitz-Thompson Estimator
- Combined Ratio Estimator
- Separate Ratio Estimator

MAS investigations led them to conclude that the Combined Ratio Estimator was the most appropriate for our requirements. Fortunately that is the method we already use anyway. The details of these options are given in a later section and the attachments.

Comparison with HMRC and NHS

Other government departments also carry out similar sample surveys to estimate fraud and error. In particular HMRC look at fraud and error in Working Families and Child Tax Credits and also NHS investigate prescription fraud and error. Although these surveys are different from our survey, they use similar estimators. Moreover, comparisons of CIs between DWP and HMRC led to some concerns amongst our stakeholders that our variances were needlessly high. So it was imperative that we compared our estimators with these alternatives to see if any of them were more appropriate or provided narrower CIs.

By applying HMRC methods to our datasets, MAS found that the standard errors were comparable to our current methods. This implied that it was the underlying data that was driving high variances rather than the method.

Formula-based CIs versus simulation

The DWP CIs are calculated using bootstrapping which is a Monte Carlo simulation technique. Although this is an established and versatile technique, it is not widely used for reporting CIs in government. We use bootstrapping for a number of reasons, but the most important are that it means we make no underlying assumptions about our population, eg. the normality assumption is not required. Secondly, there are many adjustments we make to our sample data to account for under coverage and operational issues. Trying to capture the additional uncertainty (above and beyond the usual sampling error) for each adjustment is very difficult using standard formulae-based techniques. However, bootstrapping allows us to capture these uncertainties in a natural way.

This project allowed us to compare the CIs that come from formulae-based methods and bootstrapping. By conducting this analysis we were able to refute the accusation that bootstrapping was the reason for high CIs.

Planning for Universal Credit

From 2014, there is a major change to the benefits system with the introduction of Universal Credit (UC). This is a new benefit that amalgamates various benefits currently administered by DWP and HMRC. It means that when we have to measure fraud and error in UC, we need to decide on the best estimator. This could be the DWP estimator or the HMRC estimator or some other alternative. Therefore we wanted this review to help inform our future choices for UC.

Some additional details about the analysis

In our fraud and error results we report both i) the proportion of cases overpaid and ii) the percentage of expenditure overpaid. It is the latter that is of most importance as this is used to measure Departmental performance in reducing fraud and error.

Each benefit is subject to a number of adjustments. These adjustments may be specific to the benefit or related to under-coverage in the sampling frame or administrative datasets or they could be for operational or historical reasons. Further more, different benefits can have different adjustments. To conduct analysis for this project we supplied MAS with an Income Support (IS) dataset and for ease of analysis we ignored most of the adjustments. This is because we were interested in the best approach for the underlying base sample data, excluding adjustments.

The estimators

ONS identified three options for estimating the total fraud and error in a benefit payment. These were:

The Horvitz-Thompson (HT) Estimator: $\hat{t}_{HT} = \sum_{hk} d_h y_{hk}$

The Combined Ratio Estimator: $\hat{t}_{cre} = t_X \left(\sum_{hk} d_h y_{hk} \Big/ \sum_{hk} d_h x_{hk} \right)$

The Separate Ratio Estimator: $\hat{t}_{sre} = \sum_{h} t_{Xh} \left(\sum_{k} y_{hk} \Big/ \sum_{k} x_{hk} \right)$

Where: y_{hk} is the measured amount of fraud of error for the kth case from the hth stratum, d_h is the design weight for that stratum and x_{hk} the auxiliary variable for that case which is the amount of benefit paid to the case. The t_{Xh} are total costs of the benefit for stratum h and t_X their sum over strata.

The above were considerable simplifications from how the estimates are or would be applied in practice as a number of adjustments are made. For example, the auxiliary totals include payments to cases that cannot be sampled, so include an under-coverage adjustment.

ONS Analysis of estimators

ONS went through each of the three options for the estimator in turn: Horvitz-Thompson Estimator, Combined Ratio Estimator and Separate Ratio Estimator. These estimators give an initial estimate of the total overpayment which is converted into a proportion using an extrapolated total benefit figure. This proportion is then multiplied to a forecasted total benefit figure which is more reliable.

Horvitz-Thompson Estimator

The Horvitz-Thompson (or Mean Per Unit) estimator is the simplest of the three estimators and essentially sums the weighted overpayment total.

Once an initial estimate of the overpayment has been calculated, this is divided by the extrapolated total benefit figure to give an estimated proportion of overpayment. This technique gives the following results:

Table 1: Horvitz-Thompson initial estimates and standard error

Benefit Type	**Initial estimate of overpayment**	**Standard error**	**Estimate of % overpayment**	**95% confidence interval**
IS Disabled	236,485,668.00	18,604,372.00	4.54%	(3.84%, 5.25%)
IS Lone Parent	118,614,155.52	13,616,738.00	5.24%	(4.06%, 6.42%)

Combined Ratio Estimator

The combined ratio estimator requires a little more work, but is still fairly simple. This time the weighted estimate of overpayment is summed and then divided by the weighted estimate of the total award to give a proportion of overpayment in the sample. This is then multiplied by the total annual benefit given across all local authorities to give an estimate of the total overpayment.

The rest of the notation is the same as for the Horvitz-Thompson estimator.
Once an estimate is produced for the overpayment, this is divided by the extrapolated total benefit to give a proportion of overpayment. This technique gives the following results:

Table 2: Combined Ratio initial estimates and standard errors

Benefit Type	**Initial estimate of overpayment**	**Standard error**	**Estimate of % overpayment**	**95% confidence interval**
IS Disabled	216,907,298.44	18,429,466.00	4.17%	(3.47%, 4.86%)
IS Lone Parent	110,220,894.48	13,366,800.00	4.87%	(3.71%, 6.03%)

Separate Ratio Estimator

The separate ratio estimator is the most complex of the three. For this estimator, a ratio is calculated for each stratum and then applied to the total benefit for each stratum, before it is summed to give an estimate of total overpayment. The only difference from the Combined Ratio Estimate is that it is calculated for each individual stratum h.

As before, once a final estimate is produced for the overpayment, this is divided by the extrapolated total to give a proportion of overpayment. This gave the following results:

Table 3: Separate Ratio initial estimates and standard errors

Benefit Type	**Initial estimate of overpayment**	**Standard error**	**Estimate of % overpayment**	**95% confidence interval**
IS Disabled	216,117,039.66	18,129,145.00	4.15%	(3.47%, 4.84%)
IS Lone Parent	103,604,237.54	12,937,462.00	4.58%	(3.46%, 5.70%)

Since some local authority groups are quite small, which leads to concerns about bias in the estimate itself, MAS calculated separate ratios for the regions (previously Government Office Regions, GORs) and got the following results.

Table 4: Separate Ratio initial estimates and standard errors using regions

Benefit Type	**Initial estimate of overpayment**	**Standard error**	**Estimate of % overpayment**	**95% confidence interval**
IS Disabled	217,799,084.92	18,383,462.00	4.19%	(3.49%, 4.88%)
IS Lone Parent	108,901,232.00	13,365,606.00	4.81%	(3.65%, 5.97%)

There appears to be little difference between the standard errors for the separate ratio estimators for each region and the combined ratio estimator.

ONS Conclusions on the best estimator

The ONS investigation looked into the three types of estimator, Horvitz-Thompson, Combined Ratio and Separate Ratio estimators. Looking at tables 1 to 4, the standard errors using the separate ratio estimator are lowest for the two benefits, however there is a danger that the standard errors are artificially small as some strata, on which the ratios are based, are quite small and many will have had no fraud or error observed.

Moreover, the separate ratio estimator for regions does not account for much more variance than the combined ratio estimator, so will not be considered.
The Horvitz-Thompson estimates are much larger than those calculated currently, 4.54% vs. 4.17% and 5.24% vs. 4.87%. The standard error is also larger than the other methods. This may cause problems as the figures are compared over time.

That left the combined ratio estimator. This gave a standard error which could be trusted, unlike the separate ratio estimator, and the estimate is similar to what was calculated before.

The ONS concluded that the Combined Ratio Estimator was probably the most appropriate choice. This was reassuring to us as it is that one that we currently use. However ONS also pointed out that there was a downside to this estimator in how we apply a particular adjustment regarding the combining of official error and claimant error/fraud separately. The full report by ONS here is attached here:

\\dfz72670\
101541001\Workgrou

Comparison with HMRC

We also asked ONS to compare the Combined Ratio Estimator, which DWP use, with the HMRC estimator. HMRC also measure fraud and error in tax credits every year; however their method to estimate total benefit fraud and error is different from that used by DWP. As well as using different estimators we use different methods to ascertain confidence limits. DWP uses bootstrapping, whereas HMRC uses an analytical method. The DWP estimator is known commonly as the Combined Ratio Estimator (CRE) and the HMRC estimator is a type of Horvitz-Thompson Estimator (HT). Before Universal Credit comes in October 2013, it is important to have a common estimator.

ONS used IS (Disabled and Lone Parent) data from DWP to compare the current DWP method with the HMRC method. Essentially the HMRC method is the same as the Horvitz-Thompson estimator, except that there are no grossing factors or adjustments used. The HMRC method looks like it was primarily set up to estimate the average size of fraud and/or error per case rather than the total fraud/error.

Although the HMRC method is essentially the Horvitz-Thomspon method, there is a slight difference in how the confidence intervals are calculated. The HMRC use a Satterthwaite approximation to estimate the number of degrees of freedom, but this has very little overall effect. It changes the confidence intervals by about 0.01 percentage points. This is because the degrees of freedom are quite high, so the t-values used are very close to the normal values in the confidence interval calculation.

The ONS analysis showed there was very little difference in the magnitude of the confidence intervals, so it is not the method which is driving the difference. This leads to the conclusion that the discrepancy may be due to the underlying DWP data being different to the HMRC data. The ONS report on the comparison is given here:

\\Dfz72670\
101541001\Workgrou

Comparison with Bootstrapping

The final potential source of large CIs, compared with HMRC, could be due to DWP using bootstrapping whereas ONS used formulae-based methods. We compared table 3, the combined ratio estimate with our published figures. The spread around the estimate for IS D is 17% either side of the estimate. For IS LP, it is 24% either side of the estimate. We checked these results with our previously published estimates for the same period, the spreads were very comparable. So it is interesting that our bootstrapping results are comparable with the analytical methods.

Final Remarks

In this paper we describe the reasons we applied for a QIF bid looking for advice and guidance on calculating central estimates and confidence intervals. We asked ONS to review the potential choices of estimators and decide on the most appropriate one. We also asked ONS to compare our current methods with HMRC methods. This was because we had larger CIs than HMRC, however we did not know if this was because of:

- our choice of estimator,
- the use of bootstrapping,
- that fact we add additional uncertainty into CIs due to adjustments, or
- because of the underlying DWP data being intrinsically different to HMRC data.

The ONS analysis revealed that our current choice of estimator (Combined Ratio Estimator) was probably the best one. The analysis showed that different estimators had similar magnitudes of CIs, but the other two choices had serious drawbacks. ONS also applied HMRC methods to our data to compare resulting confidence intervals. They concluded that the methods had no real effect on variances so it is likely that the underlying data is causing the differences. Because the estimate is effectively based on relatively few informative cases, the confidence intervals will be large.

Finally we compared the ONS formulae-based confidence intervals with those obtained from bootstrapping. This revealed that confidence intervals between the two methods were comparable and not leading to additional uncertainty. One issue that ONS did mention was that the confidence intervals need to reflect the additional uncertainty caused by each adjustment.

By Kav Aujla –September 2013