An Investigation of Outlier Treatment in the AWE

Authors: Hannah Finselbach, Salah Merad, Daniel Lewis. Methodology Directorate, ONS.

Executive Summary

(1) This paper addresses the recommendation proposed by Weale (2008), in which he requested that we test that the current method employed by AWE performs as well as the method in the AEI.

(2) The AEI outlier detection method is based on growth levels between two consecutive months. We extend this method in the AWE to deal with a rotating sample. The unmatched part of the sample is obtained via imputation. This method is applied to regular pay; for bonuses, the outlier detection method tested is based on levels.

(3) We considered treatment of outliers by post-stratification and weight dampening. The weight dampening method uses the same principles as Winsorisation, but here the dampening weights are obtained using heuristic methods, as optimal outlier weighting would be very difficult to determine in this context.

(4) We applied the proposed methods to real MWSS data (between 2006 and 2007) and we found that they sometimes show different patterns. To investigate which method performed best, we set up a simulation study, where the population is constructed from returned MWSS sample data.

(5) In the simulation study none of the tested methods consistently outperformed the others.

(6) One of the methods we tested is a robustification of the current AWE method, which reduces the sensitivity to revisions. As the simulation study showed that it compared well to the other methods, we recommend that this method be adopted in the AWE.

1. Background

(7) There are two main measures of short term earnings, the Average Earnings Index (AEI), and the Average Weekly Earnings (AWE). Although both measures use the same data source, the Monthly Wages and Salaries Survey (MWSS), they sometimes show a different pattern of year on year growth. The AEI is the current National Statistic, and the AWE is an experimental series. There are conceptual and methodological differences between the AEI and AWE, including their outlier procedures. In AWE a simple outlier detection and treatment is currently utilised for point in time data, on the full sample. The method is automated and contrasts with the method employed within the AEI, which has a manual element, and is based on the impact on growth calculated from a "matched" sample from successive months.

(8) Weale (2008) gave a thorough review of several aspects of the AWE and identified that additional development work is needed on the outlier methodology in order to move the AWE to National Statistic status (Recommendation 2):

"AWE should not become a National Statistic with the treatment of outliers present in the experimental series until it is clear that this is superior to the alternative which has been demonstrated to be workable in the construction of the AEI (but which would need development to be applied to AWE)."

(9) Weale (2008) outlined the work to be done to identify the best method for treating outliers in the AWE, and this paper compares some candidate methods with the current methods employed in the AWE.

2. Monthly Wages and Salaries Survey

(10) MWSS is stratified by industry and size. The employment size bands are defined as

Band 1 20-99 employees,
Band 2 100-499 employees,
Band 3 500-999 employees, and
Band 4 all businesses with 1000 or more employees.

(11) This is coupled with a higher level stratification two digit SIC (division) by Public/Private sector. The sampling fractions in the employment bands are 0.04 in band 1, 0.15 in band 2, 0.46 in band 3 and band 4 are fully enumerated. A unit selected for the MWSS is kept in the survey for 60 months, and then moves out, and is replaced by a new unit. Therefore there is a sample rotation of about 2% each month.

(12) Note that businesses with fewer than 20 employees are not sampled. The AWE makes an adjustment for these non-sampled smaller businesses, but in the methods compared in this report we do not consider these adjustments.

3. Current AWE and AEI outlier methods

(13) We refer to Duff (2007) for a detailed comparison of all methods in AWE and AEI. Here, we briefly outline the differences in their outlier treatment methods.

(14) Currently, in the AWE values are identified as outliers if they lie more than 6 standard deviations away from the stratum mean, with both the stratum mean and standard deviation evaluated excluding the firm in question. As new data come in, both the stratum mean and standard deviation in the relevant stratum change, so that the outlier status can change. This can lead to substantial revisions. Outliers in regular pay (total pay excluding bonuses) are detected using regular pay excluding bonuses per employee (payxbpe). For bonuses, the total pay per employee (paype) variable is used. For both regular pay and bonuses, outliers are treated using post-stratification into a fully enumerated post-stratum with the appropriate adjustments to the original stratum weights.

(15) In the AEI, the growth between two consecutive months is calculated from matched pairs, hence when the growth rate is calculated between month *t-1* and *t*, each business is included in the sample for month *t* only if it is also present for month *t-1*. A business whose reported wage per employee has changed by a factor of ten or more from one month to the next is identified as an outlier, and any changes by a factor of three or more is investigated manually by the statistician. In addition, after estimates have been produced, any firm that contributed more than +/- 0.05 per cent per annum to the movement of the overall index is treated as an outlier. This is technically an output editing procedure which could also be used in the AWE.

4. Proposed methods

(16) There are several possible methods that could be implemented within the AWE framework. We describe them below, and compare each with the current AWE methods in the following sections.

4.1 Detection of outliers based on growth

(17) One recommendation from Martin Weale's report on the AWE/AEI was that, before becoming a National Statistic, ONS should investigate the possibility of detecting outliers based on the growth of average earnings, instead of its current method based on levels of average earnings at a point in time. To calculate a growth for each unit in the sample data, we would have to impute for the non-matched sample (those units that were not sampled or did not respond in the previous period) such that it could be applied within the AWE estimation

framework. Therefore, we would be using the whole data set, not just the matched sample (those units that are in successive periods).

(18) One potential method of imputation that we could employ to create values for non-matched units is nearest neighbour (donor) imputation. For any unit that is sampled in time period *t,* but was not sampled in the previous month *t-1* we assume that the employment is the same in successive months. We used a statistical package (BANFF) to impute a value, using the responses from another unit from time period *t*-1, within the same stratum, and with the closest employment. We refer to BANFF (2003) for further details on the donor imputation methodology used in the package.

(19) Another method we investigated is imputing the stratum mean to create values for the unmatched sample.

(20) Once a matched sample has been created between all units in time period *t* we can calculate the growth as follows

$$G_i = \frac{R_i^t}{R_i^{t-1}} = \text{growth of business } i,$$

where R_i is the pay (excluding bonuses) per employee, defined as

$$R_i = \frac{y_i}{x_i},$$

where y_i is the returned pay (excluding bonuses), and x_i is the returned employment for unit i.

(21) For each stratum it is possible to calculate the mean and standard deviation of the growth, and we obtain thresholds, or cut offs k as six standard deviations from the stratum mean, where the mean and standard deviation of the growth are calculated excluding the top and bottom 1% of observations. This is to make the method robust to extreme values.

4.1.1 Treatment of outliers by weight dampening

(22) Next we define the dampening weight d_i as

$$d_i = \frac{G_i^*}{G_i},$$

where

$$G_i^* = \begin{cases} G_i & \text{if } G_i < \text{the cut off } k \\ k_i & \text{if } G_i > \text{the cut off } k \end{cases}$$

(23) These weights will be applied to the pay excluding bonuses and bonus paid responses to treat outliers. This treatment uses the same principles as Winsorisation, but here the dampening weights are obtained using heuristic methods, as optimal outlier weighting would be very difficult to determine in this context.

(24) We let

$$y^*_i = d_i * y_i$$

and then our estimate of regular pay, for example, is calculated as

$$\hat{Y} = \sum_{s} y_i + \sum_{s} (w_i - 1) y_i^*.$$

(25) We note here that outlier detection methods based on growth will not be practical for pay including bonuses, due to the volatility in the series of bonus pay.

4.2 Detection of outliers using a robust threshold

(26) Units are treated as outliers if they lay more than 6 standard deviations from the stratum mean, where the stratum mean and standard deviation are calculated excluding the top and bottom 1% of observations (removing extreme values). This should prevent late responses causing large revisions - if late responses are outliers they may still change the status of units within the stratum, but their effect on the estimates should be reduced compared to the current method.

(27) For bonuses, outliers are currently detected using total pay per employee. In addition to this, we investigate the using bonus per employee.

(28) Once outliers have been detected using the robust gates method, treatment could be either via post stratification, or weight dampening, using the same approach as described for growth in section 4.1.1. We compare both in sections 5 and 6.

4.3 Separate Winsorisation of pay variable and employment

(29) Winsorisation is used in most ONS surveys to detect and treat outliers. Winsorisation is a method for detecting and treating outliers in linear weighted estimates in a way that approximately minimises mean squared error (MSE).

(30) Neither the AWE formula for average pay in a particular period, nor the formula for the growth in average pay between two periods are linear estimates. However, it is possible to implement Winsorisation on the four 'level' estimates that make up the AWE estimate of growth and take the ratios of these Winsorised estimates.

(31) We refer to Lewis (2006) for more information on the theory of Winsorisation.

4.4 Detection of outliers in bonuses based on the ratio of bonuses paid to regular pay

(32) An alternative method to detect outliers for bonuses is to look for unusually high ratios of bonuses paid over regular pay. For each unit we calculate

$$r_i = \frac{b_i}{y_i},$$

where y_i is regular pay (excluding bonuses) and b_i is the bonuses paid by unit i.

(33) Then, if r_i is more than some threshold k the bonus paid by unit i is identified as an outlier, and treated via post stratification. Thresholds were chosen based on the sample data, and we gave different thresholds depending on the industry. For financial divisions (2 digit SIC 65, 66 and 67) we set the threshold at 4, and for all other industries the threshold was 10.

4.5 Combinations of outlier detection and treatment methods

(34) For clarity, and later notation, we list a comprehensive list of all outlier methods compared.

Average regular pay

REG METHOD 1: Detection based on growth using the stratum mean for "missing" or non-matched units. Treatment via outlier weights derived from the upper and lower thresholds.

REG METHOD 2: Detection based on growth using nearest neighbour imputation for non-matched units. Treatment via outlier weights derived from the upper and lower thresholds.

REG METHOD 3: Detection based on trimmed, robust thresholds. Treatment via post stratification.

REG METHOD 4: Detection based on trimmed, robust thresholds. Treatment via outlier weights derived from the robust thresholds.

REG METHOD 5: Detection and treatment via Winsorisation as normally implemented in ONS, but applied to regular pay and employment separately.

Average bonus

(35) In the current AWE methodology, employment is weighted using the same grossing factor as regular pay. In order to fairly compare the outlier treatment methods for average bonus paid, we fix the treatment of regular pay as outlined in REG METHOD 3 described above, and first estimate employment using the derived weights. Total bonuses paid were then estimated, and the following outlier methods applied.

BON METHOD 1: Detection based on trimmed distribution of total pay per employee, robust thresholds. Treatment via post stratification.

BON METHOD 2: Detection based on trimmed distribution of total pay per employee, robust thresholds. Treatment via outlier weights derived from the robust thresholds.

BON METHOD 3: Detection based on trimmed distribution of bonus per employee, robust thresholds. Treatment via post stratification.

BON METHOD 4: Detection based on trimmed distribution of bonus per employee, robust thresholds. Treatment via outlier weights derived from the robust thresholds.

BON METHOD 5: Detection based on ratio of bonus over regular pay. Treatment via post stratification.

BON METHOD 6: Detection and treatment via Winsorisation as normally implemented in ONS, but applied to bonus and employment separately.

5. Comparisons with current AWE method for MWSS samples

(36) An initial comparison of the alternative methods was done using MWSS survey returns (with around 9000 sample units per month) from June 2006 (month 1) to June 2008 (month 25), and calculating the estimates of levels of average regular pay and bonuses. We then calculated an estimate of the year on year growth. This is not what would be obtained in practise, as we made no adjustments for the small businesses with less than 20 employees. However, it gives an indication of the results.

(37) For simplicity estimates are compared at the whole economy level. Figures 1 to 4 compare estimates of year on year growth in regular pay (excluding bonuses) obtained from some of the proposed methods to the current AWE method. The two methods based on growth show very similar results, and so we omit the comparisons for the method based on using donor imputation to create the matched pairs, and instead show

it in Appendix A1. Comparisons of the level estimates of average regular pay are shown for interest in Appendix A2.

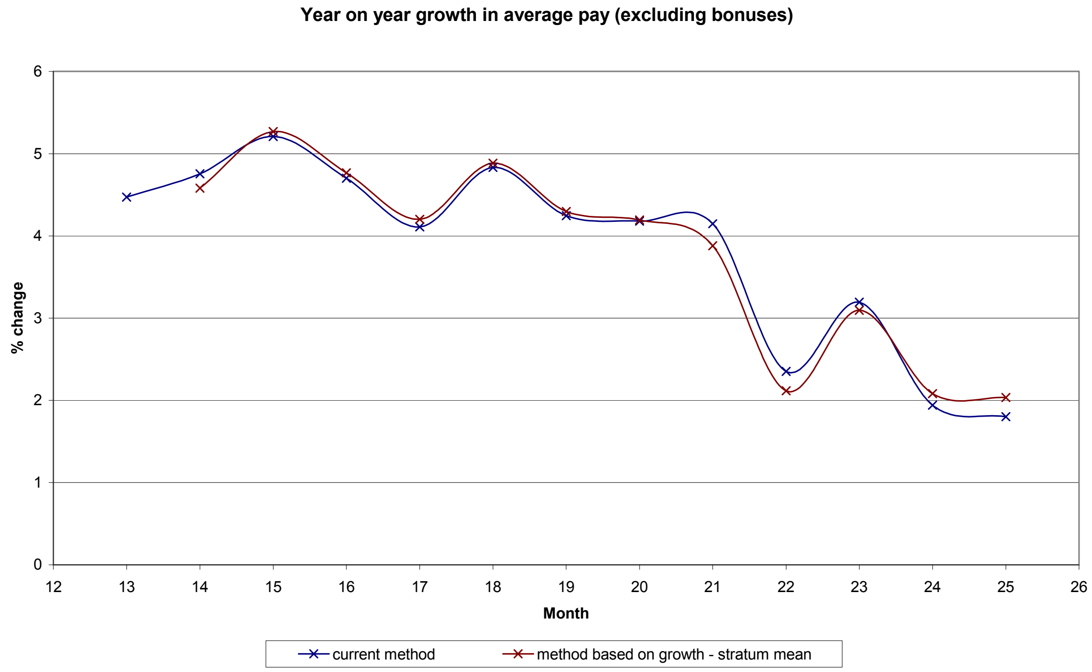

Figure 1: Estimates of year on year growth of average regular pay using REG METHOD 1 with the current AWE outlier method

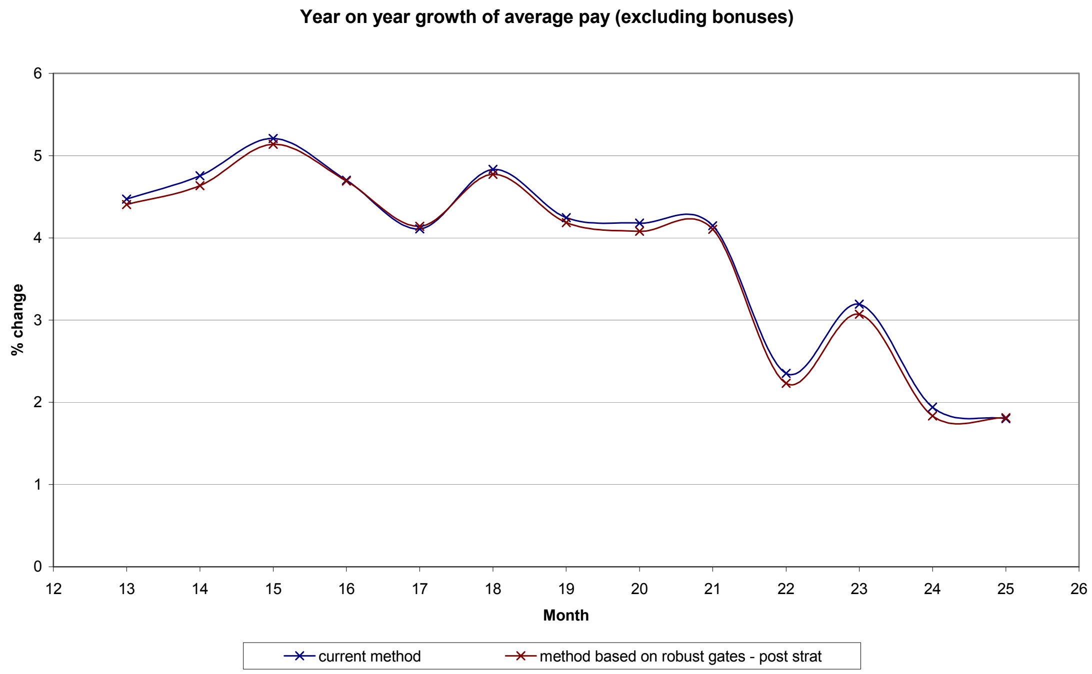

Figure 2: Estimates of year on year growth of average regular pay using REG METHOD 3 with the current AWE outlier method

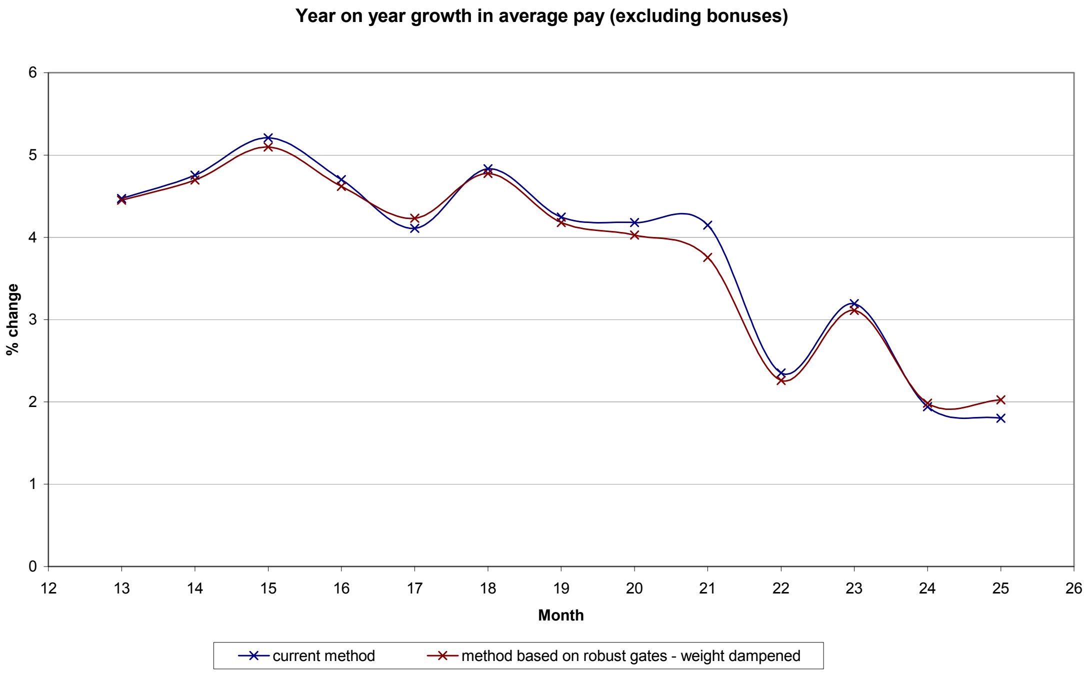

Figure 3: Estimates of year on year growth of average regular pay using REG METHOD 4 with the current AWE outlier method

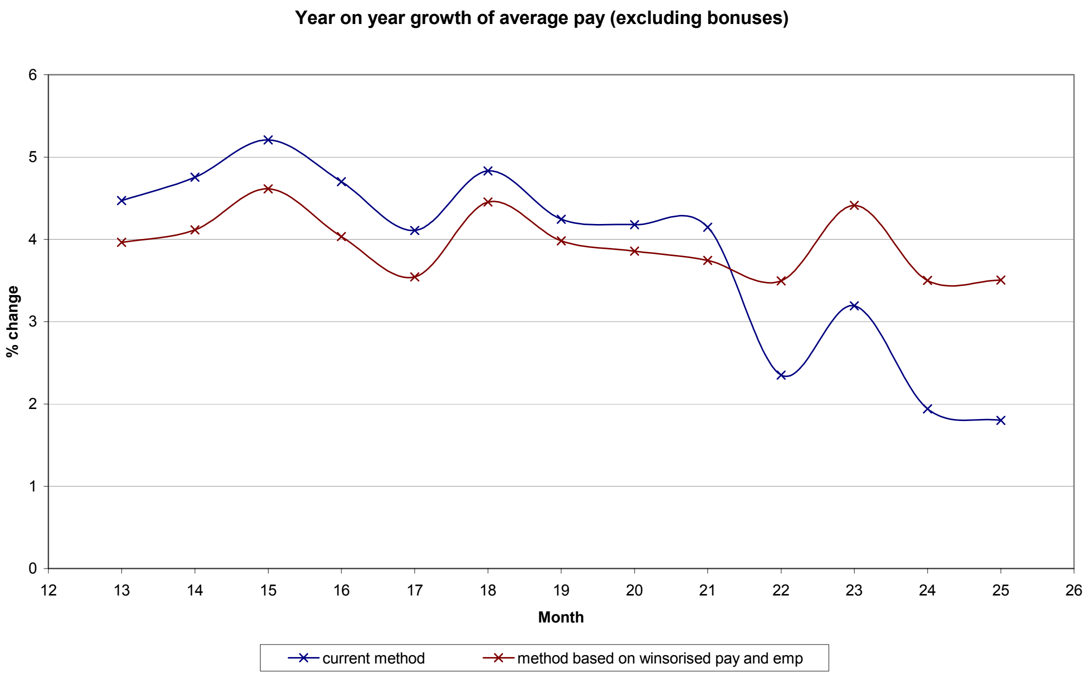

Figure 4: Estimates of year on year growth of average regular pay using REG METHOD 5 with the current AWE outlier method

(38) It is clear that for some periods the proposed methods can lead to different estimates of growth to those under the current AWE method, particularly the REG METHOD 5. Figures 5 to 8 compare estimates of year

on year growth in average bonus paid obtained from some of the proposed methods to the current AWE method. There was little difference in estimates when detecting outliers in bonuses using bonus per employee instead of total pay per employee, therefore we omit the comparisons of BON METHOD 3 and 4 with the current AWE method from the text, and they are instead shown in Appendix A1. Comparisons of the level estimates of average bonus paid are shown for interest in Appendix A2.

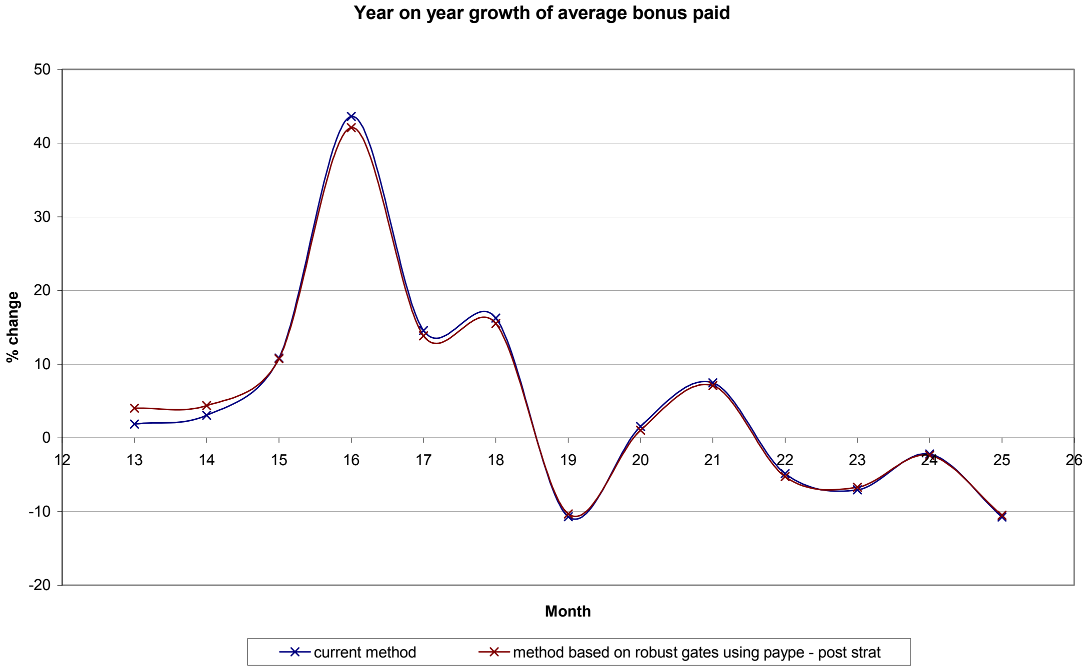

Figure 5: Estimates of year on year growth of average bonus paid using BON METHOD 1 with the current AWE outlier method

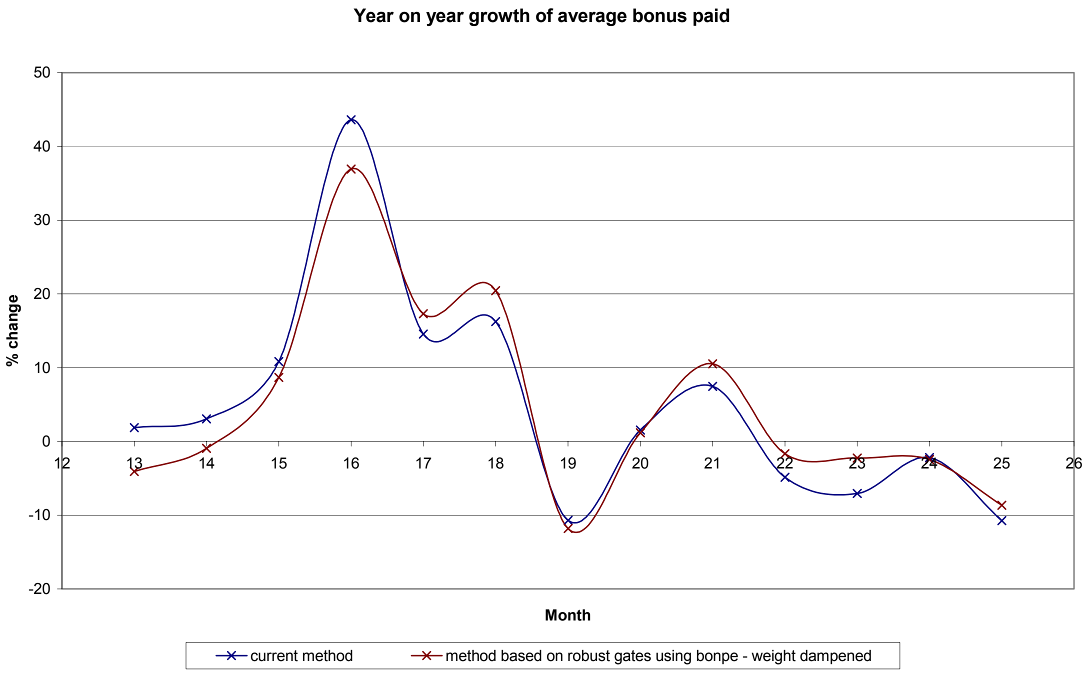

Figure 6: Estimates of year on year growth of average bonus paid using BON METHOD 2 with the current AWE outlier method

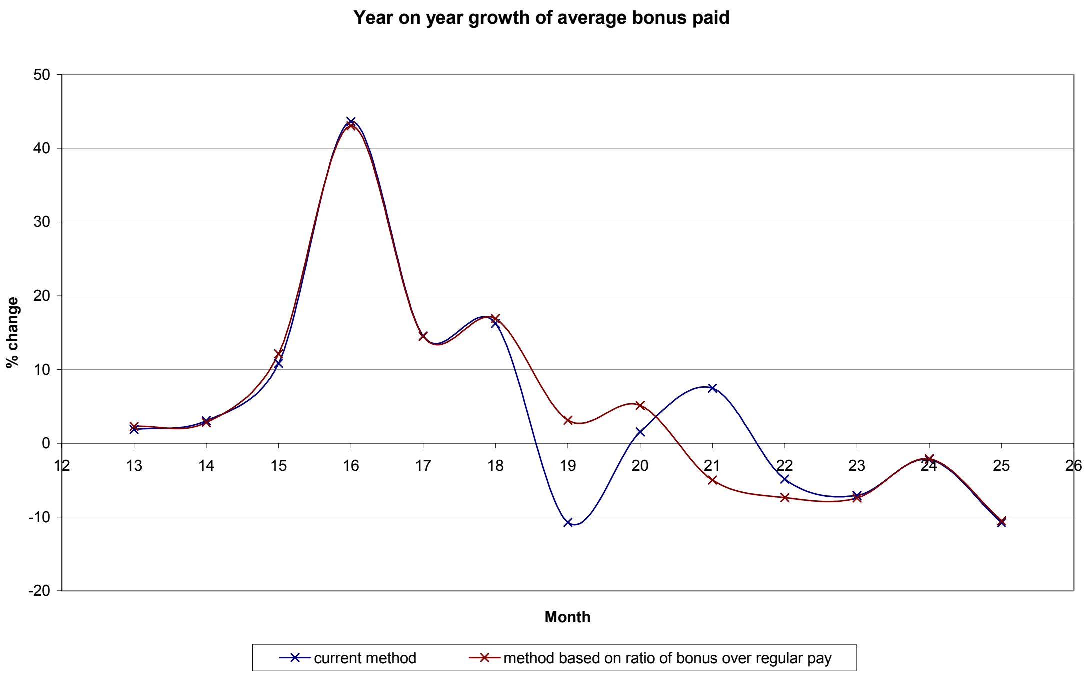

Figure 7: Estimates of year on year growth of average bonus paid using BON METHOD 5 with the current AWE outlier method

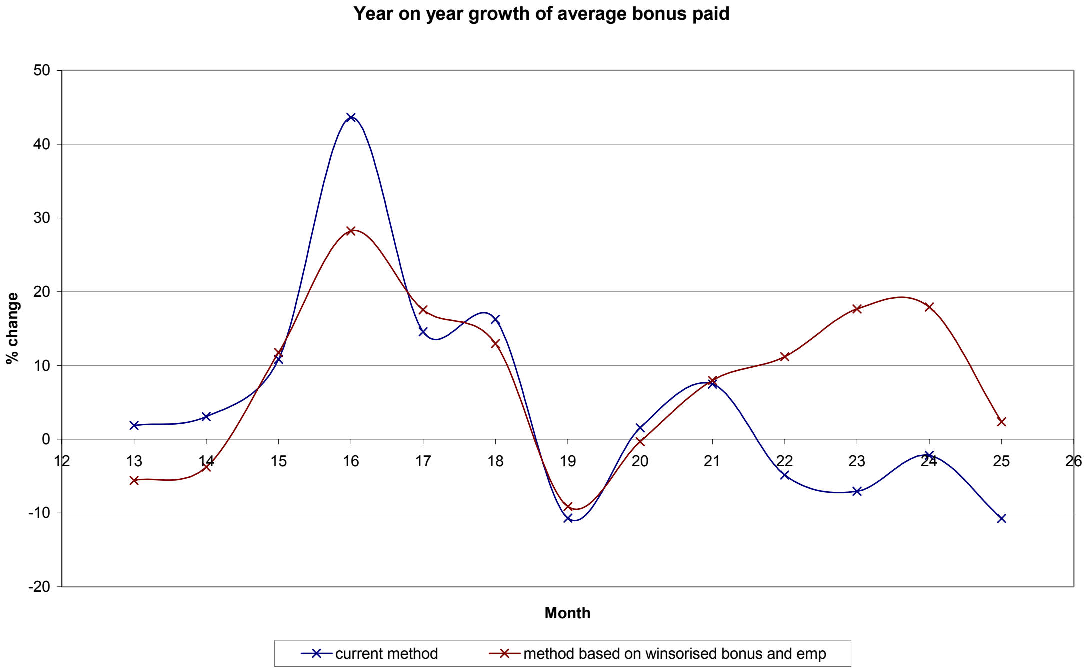

Figure 8: Estimates of year on year growth of average bonus paid using BON METHOD 6 with the current AWE outlier method

(39) Similarly to regular pay, there are some big differences between the estimates of growth in average bonus obtained using the current AWE method compared to those obtained using some of the proposed methods. To decide which method is best we need to evaluate them; we do this using a simulation study.

6. Simulation and results

(40) In order to compare the performance of our proposed methods we have conducted a simulation study, based on re-sampling from a population based on MWSS data. The population is made up of 6601 units, and nearest neighbour imputation was used to construct values for units rotated out of the original sample, therefore providing a full pseudo population for 24 periods (January 2006 to December 2007). We have excluded all fully enumerated units (those with employment greater than 1000) from the population; hence, the results will only compare the estimates obtained from sampled units. We take 2500 samples from this pseudo population. These can be compared to the known "population" parameters.

(41) In the simulation we attempted to broadly replicate the true sampling methods. Due to the small population size the strata were widened, and sampling fraction increased.

(42) The measure we used to compare the performance of the compared methods is the average relative absolute error. It is defined as

$$\text{Relative Error (\%)} = 100 \times \frac{\sum_s \left|\Delta - \hat{\Delta}_s\right|}{L\left|\Delta\right|}$$

where L is the number of samples drawn from the pseudo population. Here, $\hat{\Delta}_s$ is the estimate of Δ, which is the percentage change, or year on year growth, defined as

$$\Delta = 100 \times \left(\frac{\hat{R}^{t} - \hat{R}^{t-12}}{\hat{R}^{t-12}} \right)$$

where $\hat{R}$ is the level estimate of average earnings.

6.1 Regular pay

(43) The average relative absolute error for each method's estimate of regular pay is shown in Table 1. The method with the lowest error has been highlighted for each month. We see that there is no method that performs best every month.

Table 1: Relative error (%) of estimates of year on year growth (percent change) in average regular pay

Year on year growth	Current method	REG METHOD 1	REG METHOD 2	REG METHOD 3	REG METHOD 4	REG METHOD 5
January	24.23	.	.	22.12	26.21	38.12
February	30.56	26.42	26.47	25.96	27.22	33.72
March	23.13	23.70	23.69	24.36	22.54	24.26
April	27.66	27.85	27.79	28.22	28.81	27.16
May	28.36	27.42	27.48	27.90	28.67	25.20
June	25.69	25.37	25.39	26.24	25.27	33.25
July	32.57	30.12	30.10	31.87	31.16	40.56
August	26.25	26.14	26.21	25.96	26.63	47.08
September	30.24	26.78	26.81	29.46	27.74	29.47
October	28.81	27.90	27.93	32.00	30.13	25.90
November	23.96	25.19	25.15	29.15	28.71	23.18
December	24.55	23.93	23.93	25.56	24.85	37.52

6.2 Bonuses

(44) The average relative absolute error under each method for bonuses is shown in Table 2. Again, there is no method that is performing better for every period. The current method has the lowest relative error for some months. Winsorisation does not appear to perform as well for bonuses as it did for regular pay.

Table 2: Relative error (%) of estimates of year on year growth (percent change) in average bonus paid

Year on year growth	Current method	BON METHOD 1	BON METHOD 2	BON METHOD 3	BON METHOD 4	BON METHOD 5	BON METHOD 6
January	105.94	97.61	106.26	103.60	114.11	93.60	79.82
February	510.42	466.55	547.24	355.45	426.44	844.56	255.32
March	10728.95	8312.63	8094.49	8021.42	7878.16	6901.51	10588.63
April	136.99	140.95	128.88	140.20	124.82	137.94	309.25
May	5418.28	4697.10	5233.78	7864.07	5789.49	5498.68	9940.73
June	61.91	50.47	48.16	53.47	44.92	50.20	71.54
July	265.42	229.44	226.05	185.42	229.54	262.81	248.69
August	33.66	44.76	38.67	36.96	42.06	46.01	62.48
September	43.22	40.15	42.51	41.58	49.94	45.02	52.88
October	288.13	337.21	315.91	171.97	192.54	345.73	374.34
November	214.83	174.07	210.27	96.68	134.67	221.87	120.33
December	66.62	83.79	69.67	95.17	69.95	75.26	69.74

(45) We note that the relative errors are markedly higher than those calculated for regular pay, confirming that due to the volatile nature of bonuses, they are difficult to estimate. In particular, we notice that for months March and May the relative errors are considerably higher than other months. For these months the true growths (percentage change) calculated from the pseudo population are 0.11 and -0.5 respectively, and such small growths appear to be difficult to estimate, regardless of the outlier methodology used, which is not unexpected.

(46) NOTE: The purpose of this simulation is purely to compare methods, and neither the pseudo population growths nor the estimates are comparable to the real MWSS population and publications.

7. Recommendations

(47) The simulation study showed that estimates of growth were fairly consistent for all outlier treatment methods, and the more complicated and sophisticated methods developed were not always superior to the current method. Therefore, we recommend that the AWE continues to detect outliers based on the level of the returned pay per employee, but implement the more robust thresholds used in REG METHOD 3 and BON METHOD 1, outlined in section 4.2. Intuitively these thresholds should be less sensitive to revisions, which is a problem that has been identified in the AWE.

(48) By removing the extreme values before calculating the thresholds, if a firm replies after the provisional results have been calculated its effect on the thresholds (and therefore the status of identified outliers for the provisional data) should be reduced compared to the current method. Further research would be needed to explore the issue of outliers and revisions, as it was not the focus of this report.

8. References

BANFF (2003). Functional description of the BANFF system for edit and imputation. Statistics Canada.

Duff, H. (2007). A preliminary analysis of the differences between AWE and AEI. *Economic & Labour Market Review* **9**, pp 48-56.

Lewis, D. J. (2006) An examination of the use of Winsorisation in sample surveys. Masters Thesis, University of Southampton.

Weale, M. (2008). The Average Earnings Index and Average Weekly Earnings. *Report submitted to the National Statistician*.

APPENDIX

A1. Year on year growth of average pay (excluding bonuses) and average bonus paid for other methods compared

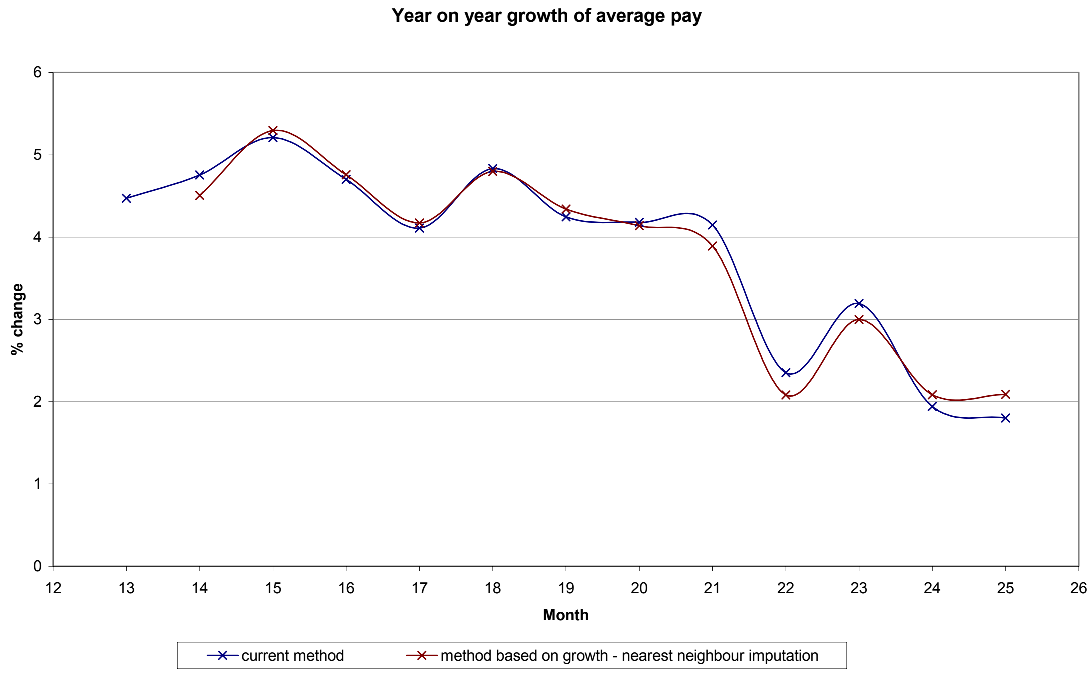

Figure A1.1: Estimates of year on year growth of average regular pay using REG METHOD 2 with the current AWE outlier method

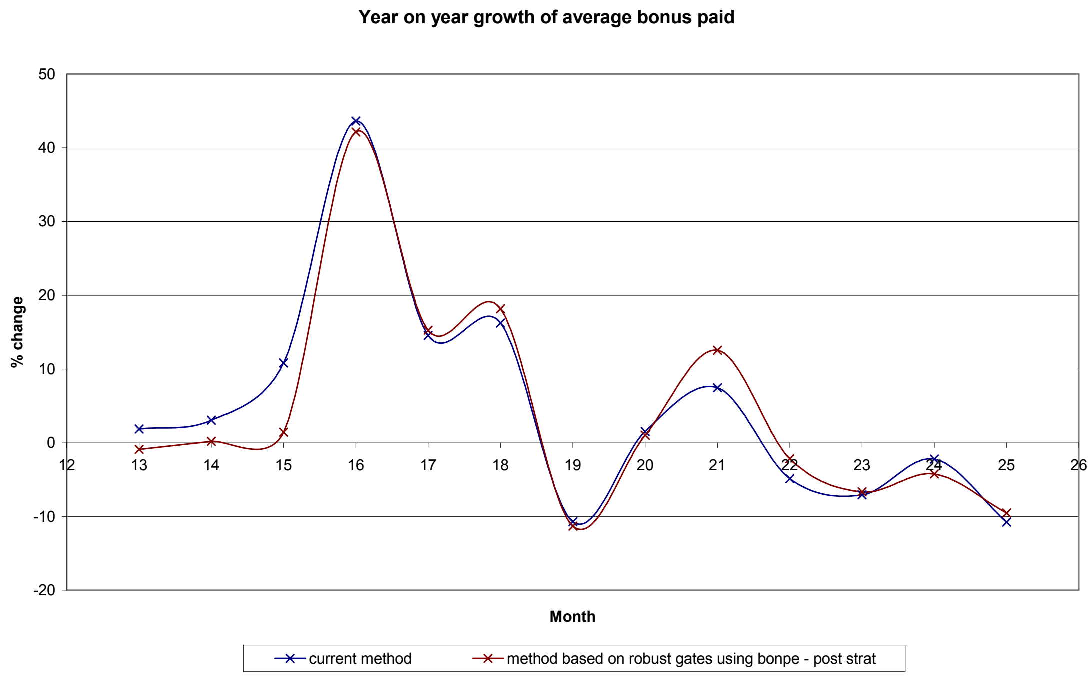

Figure A1.2: Estimates of year on year growth of average bonus paid using BON METHOD 3 with the current AWE outlier method

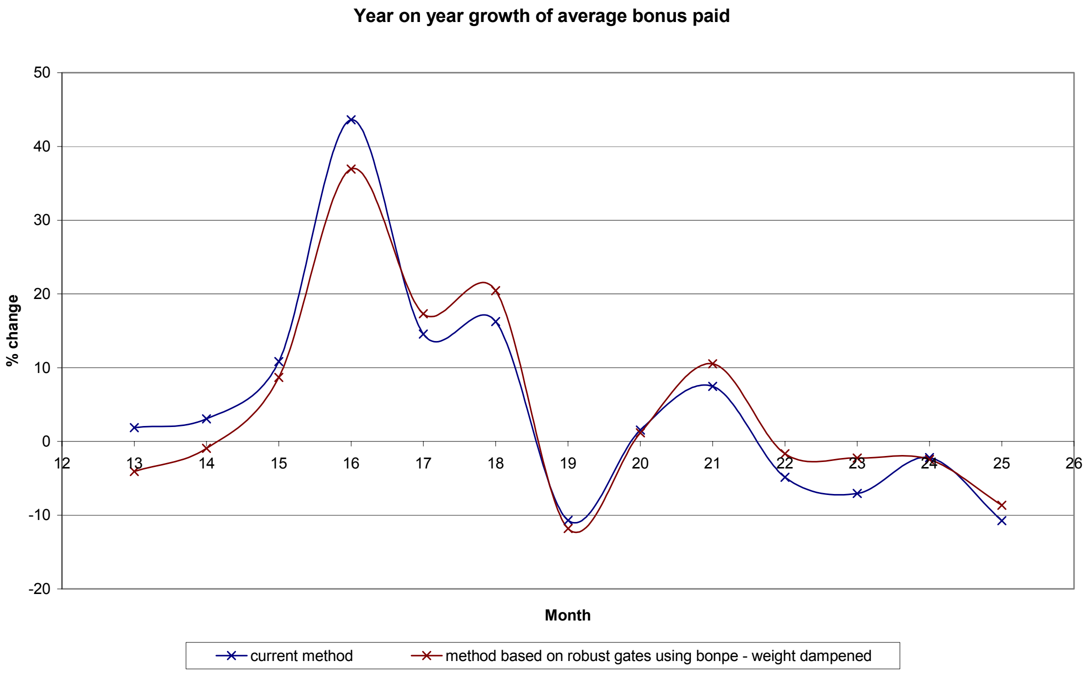

Figure A1.3: Estimates of year on year growth of average bonus paid using BON METHOD 4 with the current AWE outlier method

A2. Estimates of levels of average pay (excluding bonuses) and average bonus paid for all methods compared

Regular pay

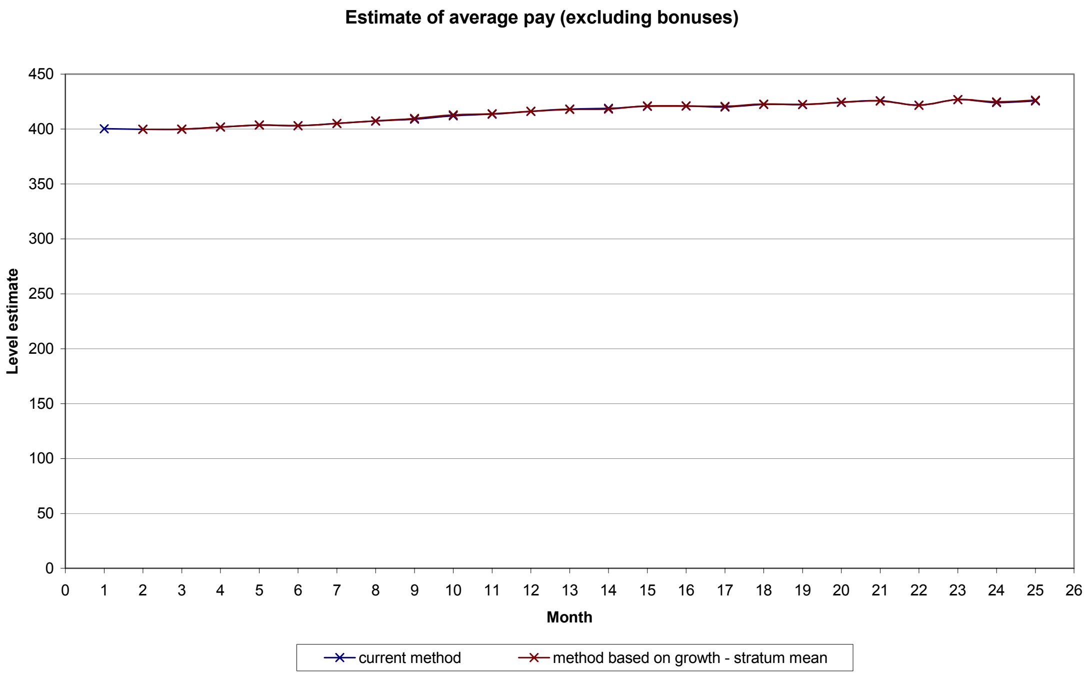

Figure A2.1: Estimates of average regular pay using REG METHOD 1 with the current AWE outlier method

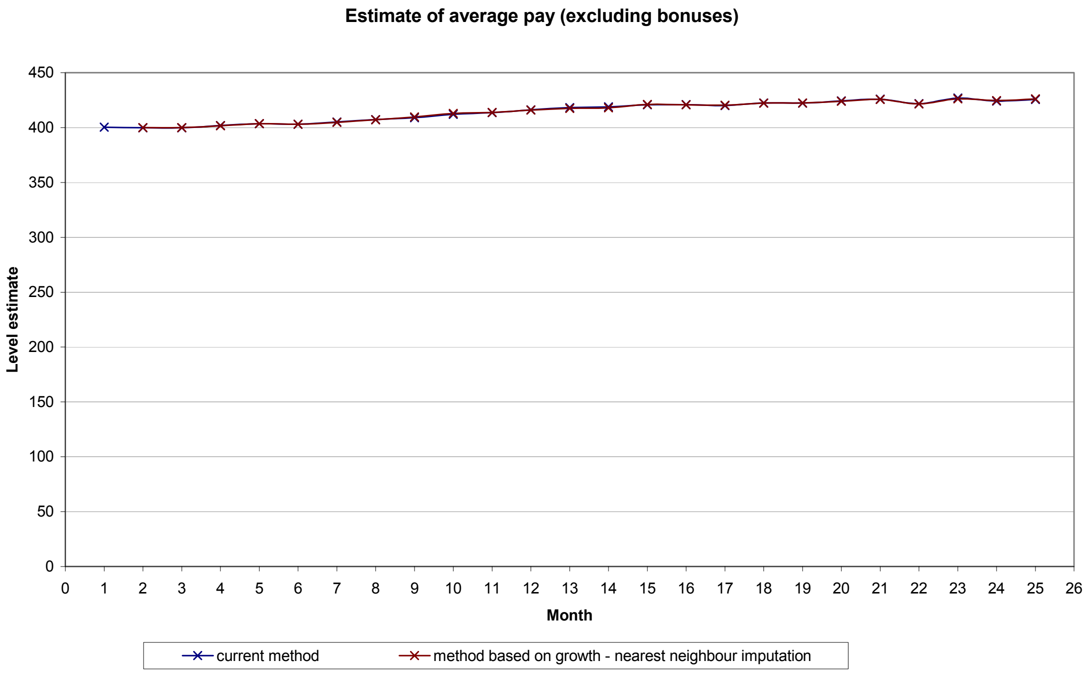

Figure A2.2: Estimates of average regular pay using REG METHOD 2 with the current AWE outlier method

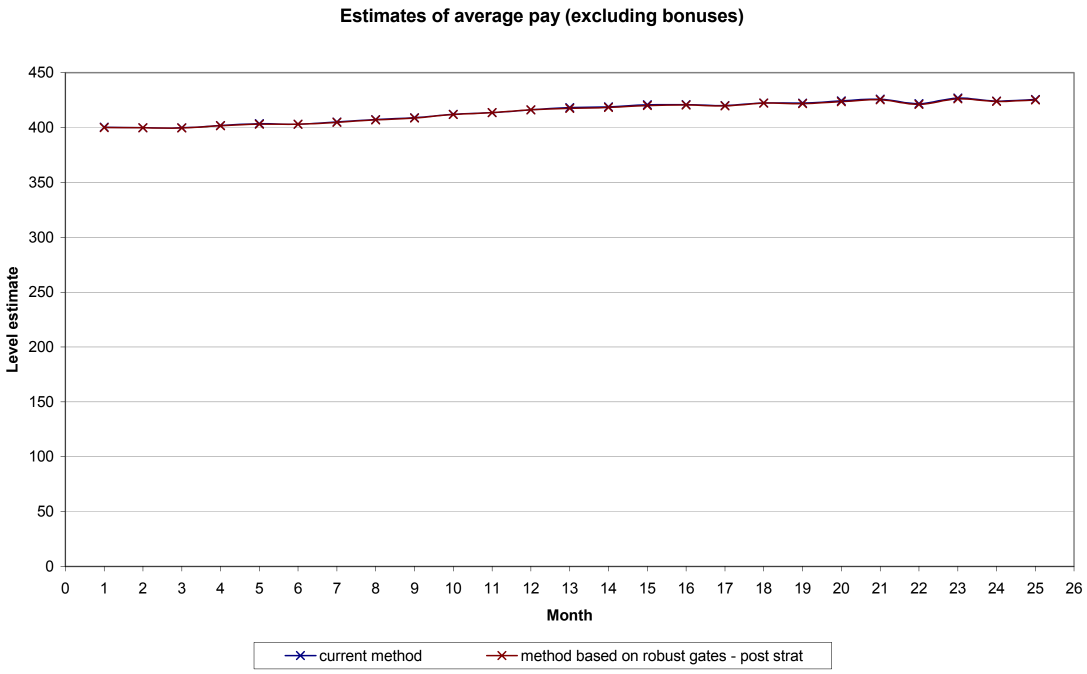

Figure A2.3: Estimates of average regular pay using REG METHOD 3 with the current AWE outlier method

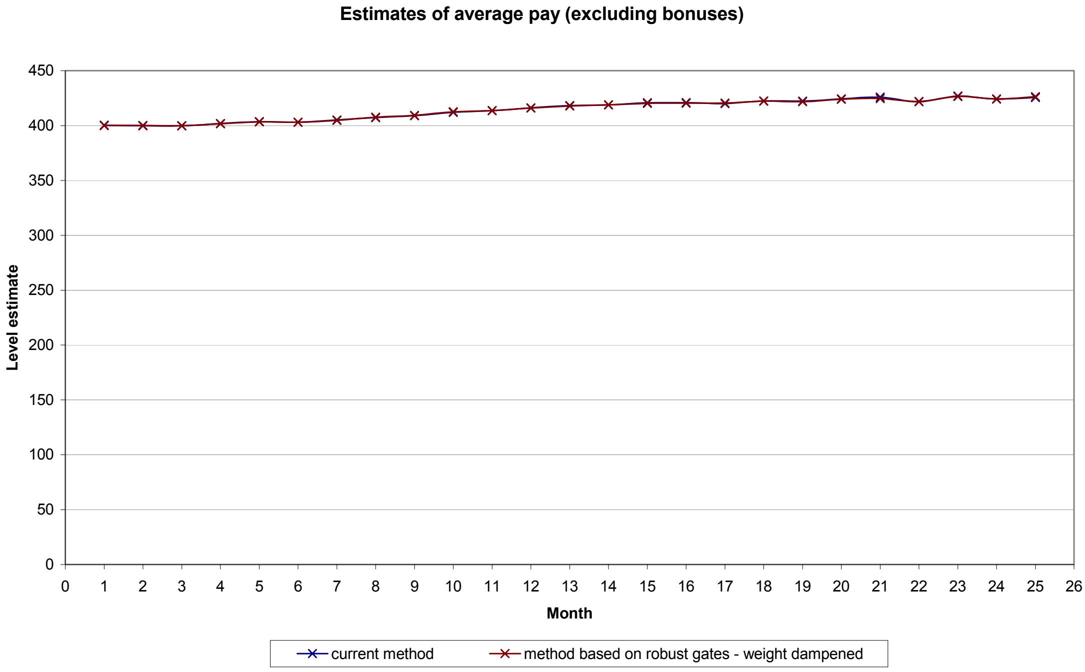

Figure A2.4: Estimates of average regular pay using REG METHOD 4 with the current AWE outlier method

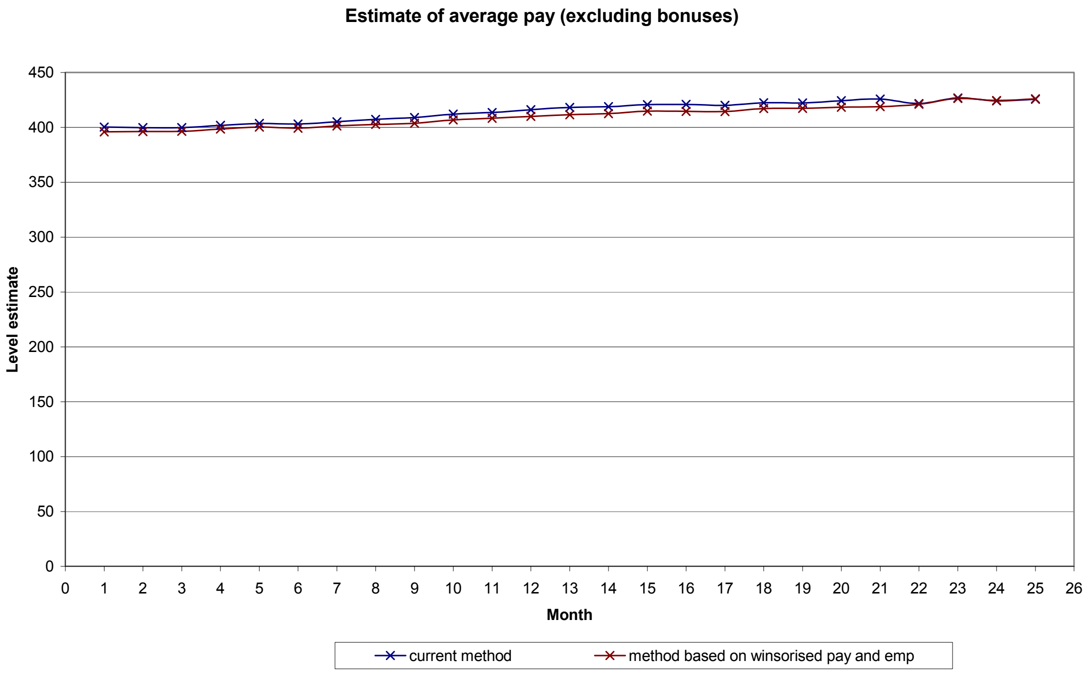

Figure A2.5: Estimates of average regular pay using REG METHOD 5 with the current AWE outlier method

Bonuses

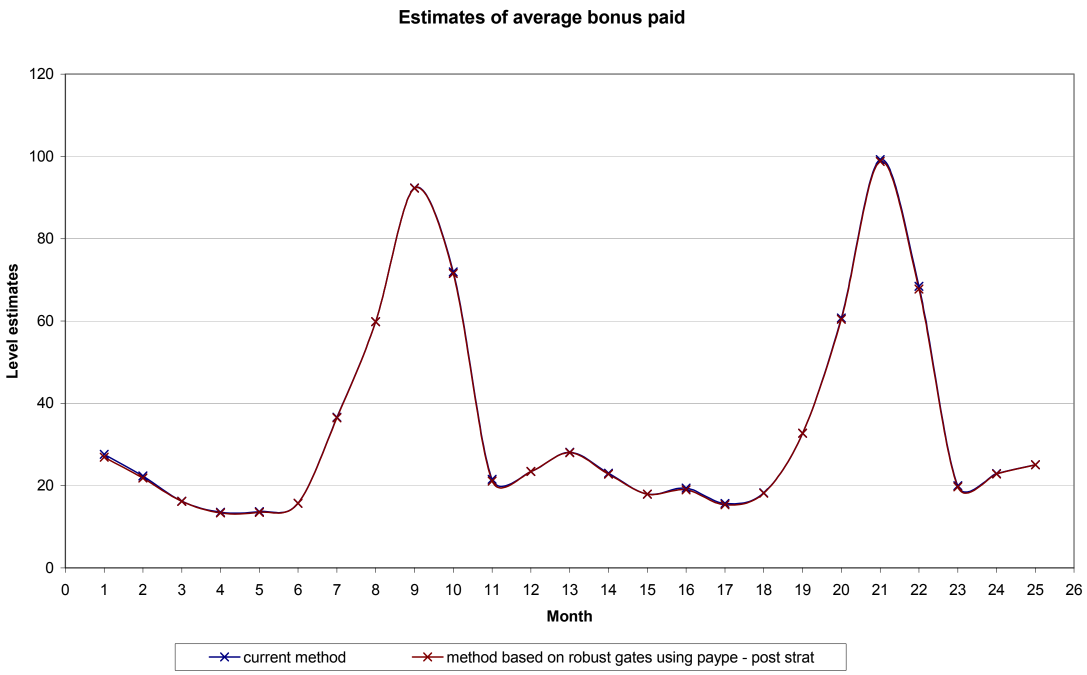

Figure A2.6: Estimates of average bonus paid using BON METHOD 1 with the current AWE outlier method

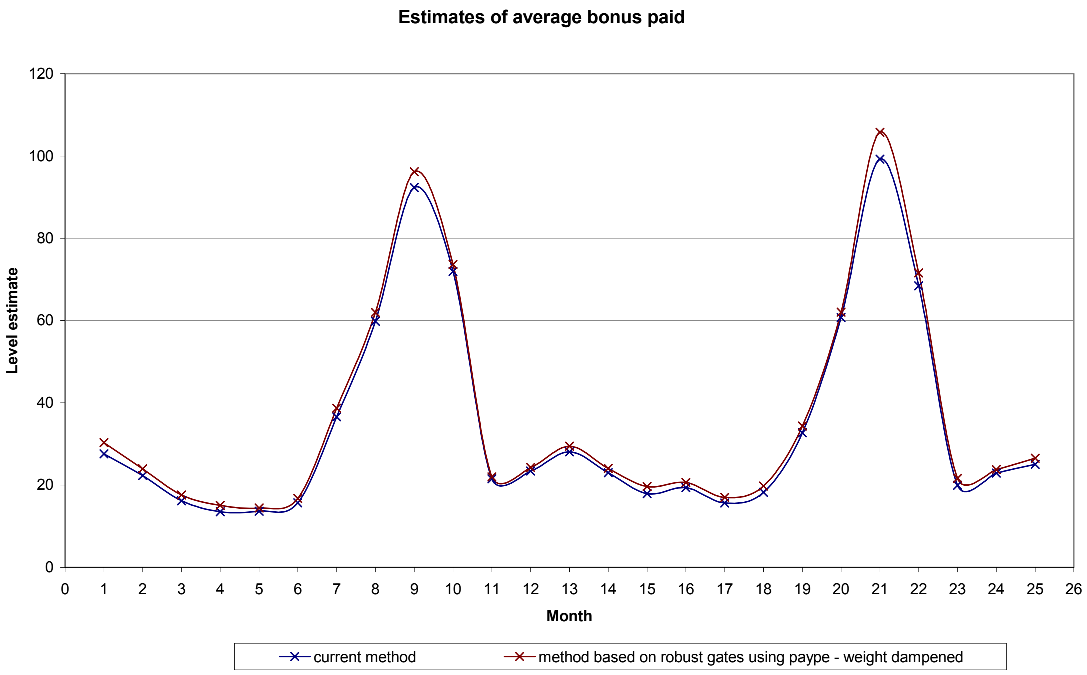

Figure A2.7: Estimates of average bonus paid using BON METHOD 2 with the current AWE outlier method

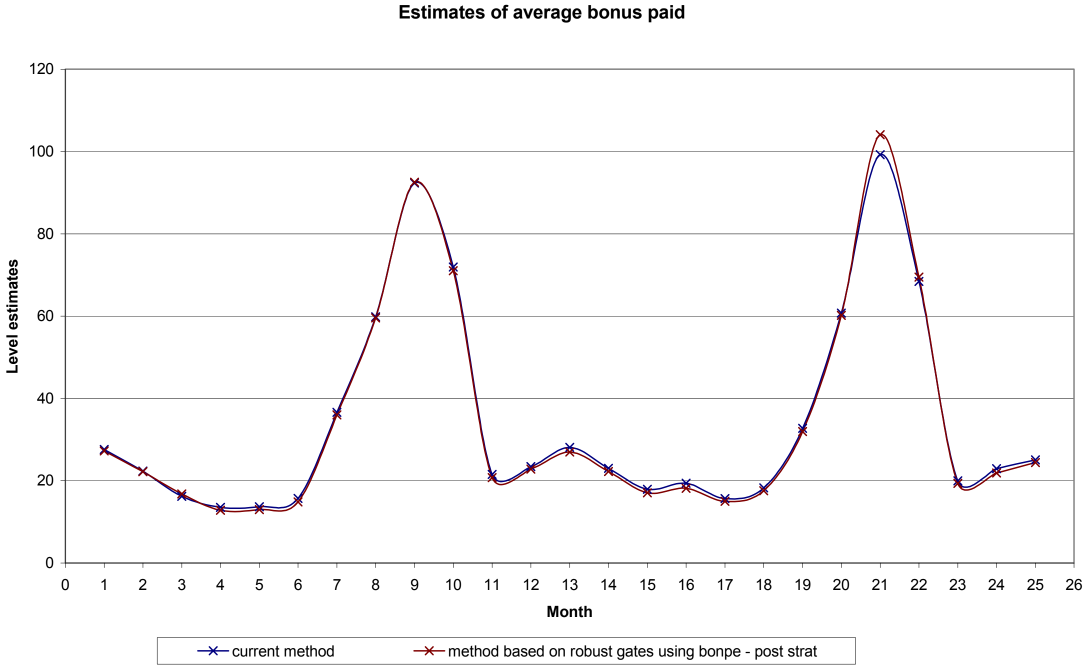

Figure A2.8: Estimates of average bonus paid using BON METHOD 3 with the current AWE outlier method

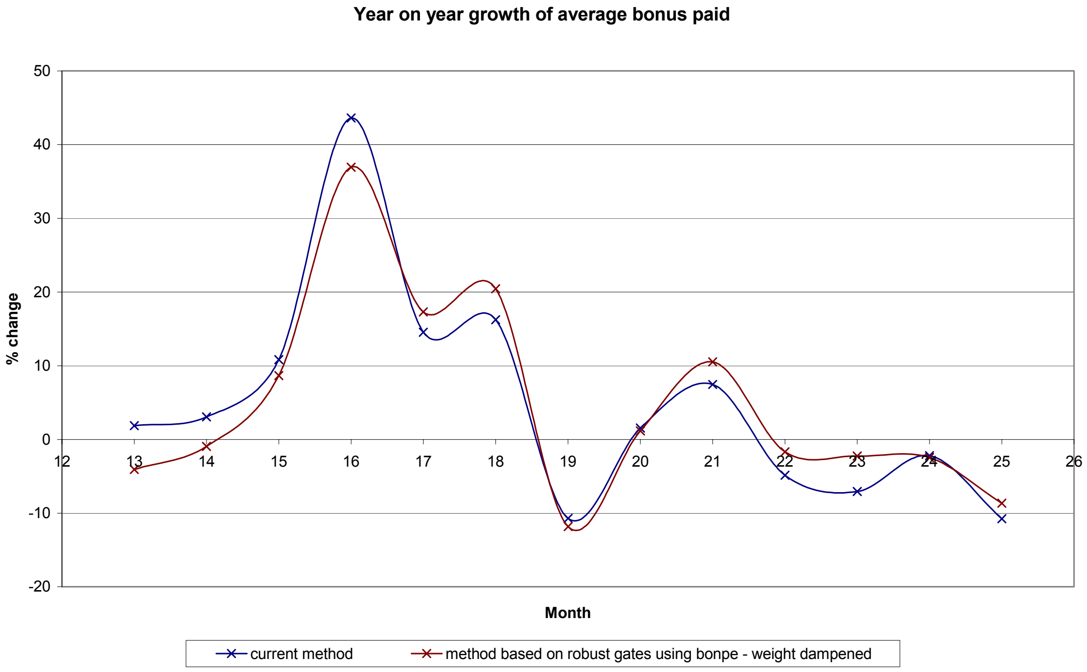

Figure A2.9: Estimates of average bonus paid using BON METHOD 4 with the current AWE outlier method

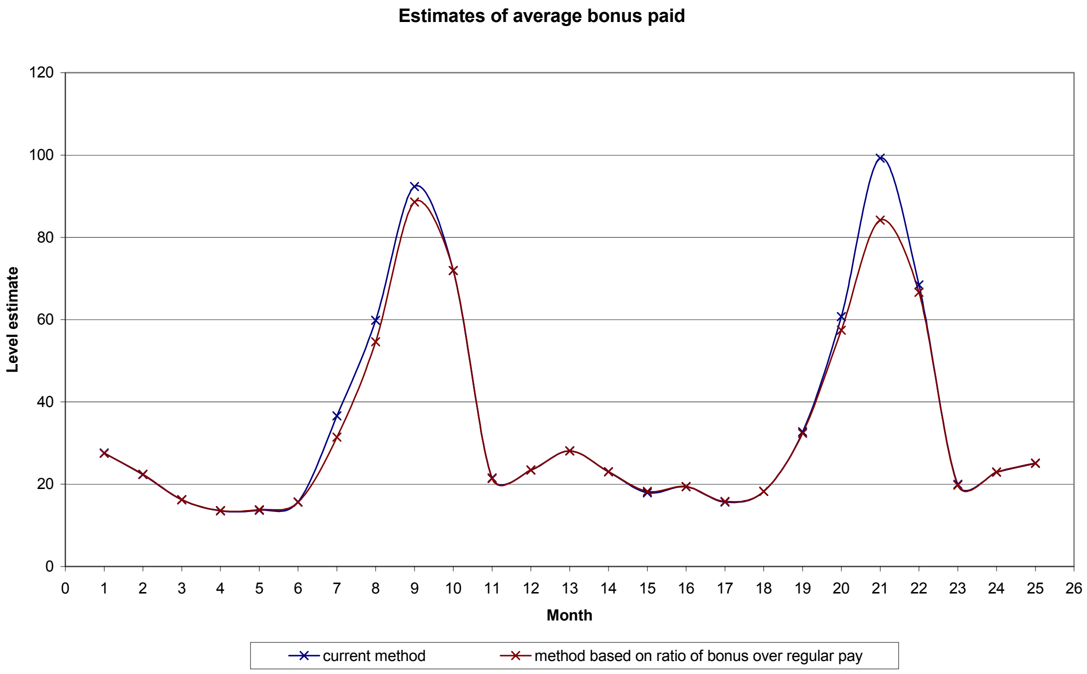

Figure A2.10: Estimates of average bonus paid using BON METHOD 5 with the current AWE outlier method

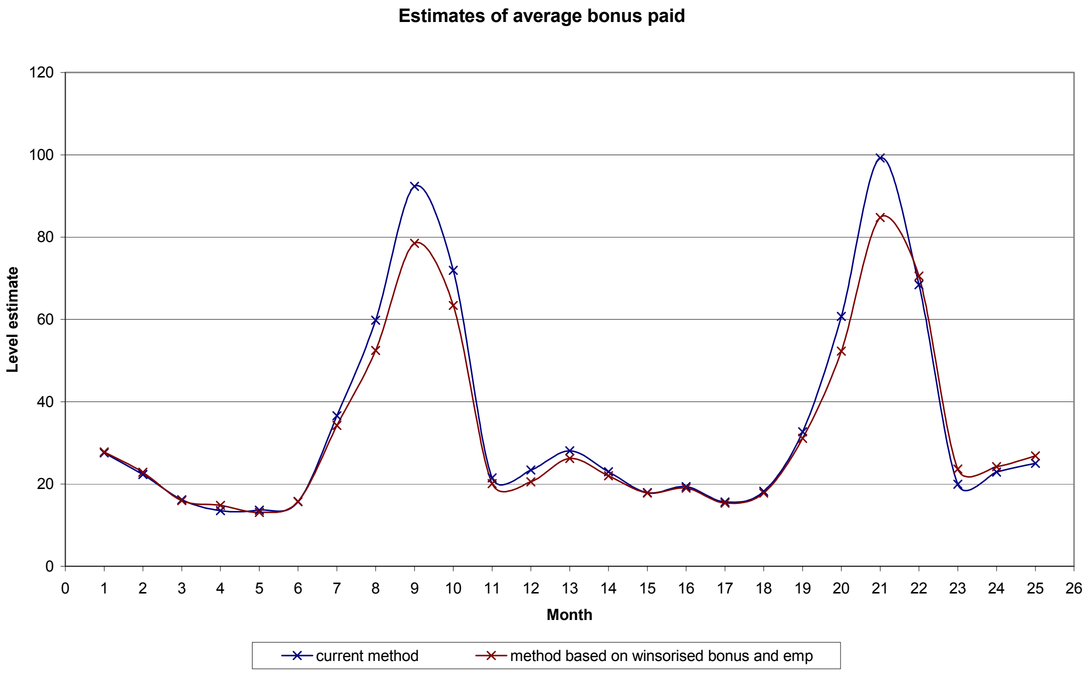

Figure A2.11: Estimates of average bonus paid using BON METHOD 6 with the current AWE outlier method