Estimates of sampling variability for Workforce Jobs

Summary

This note describes the methods used to calculate estimates of sampling variability for workforce jobs. WFJ sampling errors will be published as coefficients of variation and 95% confidence intervals and should be used as a guide rather than definitive measures of precision. They will be produced on an annual basis, taken at Q2 each year as this quarter is least affected by seasonality. This new set of sampling errors, displayed on pages four and five, supersedes those published in the 2006 employment and jobs review.

Updated methods and systems used to produce Workforce Jobs estimates have made it possible to derive approximate standard errors. Since Workforce Jobs is a composite measure drawing on a range of sources, some assumptions and approximations need to be made which are described in this note. It is recommended that this note be used alongside the Workforce Jobs Quality and Methodology information paper (QMI) which provides a detailed description of related methods and sources.

Method

Workforce Jobs (WFJ) is the sum of Employee Jobs (EJ), Self-Employment Jobs (SEJ), Government Supported Trainees (GST) and Her Majesty's Forces (HMF). Employee Jobs and Self-employment jobs are sample-based estimates and hence have associated sampling errors. The private sector element of EJ is measured primarily by the Short Term Employer Surveys (STES), which is benchmarked to the annual Business Register and Employment Survey (BRES). The public sector element comes from the Quarterly Public Sector Employment Survey (QPSES), which is a quarterly census, while GST and HMF jobs are sourced from non-sample based administrative sources. Thus sampling variability for EJ is specific to the private sector surveys. It is assumed that the two components EJ and SEJ are independent, so their respective variances are computed separately and added together to create a total estimate of variance for Workforce Jobs.

Thus:

$$S_{WFJ} = \sqrt{S^2{}_{EJ} + S^2{}_{SEJ}} \qquad \text{(eq.1)}$$

where S is the standard error.

Employee Jobs component

The STES estimator is defined and described on pages four to six of the WFJ QMI. Variances for STES are calculated in the Generalised Estimation System (GES) based on a one-stage cluster model with

element level calibration. The variance for the STES-estimated employment $\hat{Y}$ for a given domain d, where i is a cluster (reporting unit) and k is an element (local unit) is defined as:

$$V\hat{a}r(\hat{Y}_d) = \sum_h \frac{N_h - n_h}{N_h} \frac{n_h}{n_h - 1} \sum_{i \in s_h} (u_{hi} - \bar{u}_h)^2 \qquad \text{(eq. 2)}$$

where $\quad u_{hi} = \sum_{k \in i} \frac{N_h}{n_h} g_k e_{[Y]dk} \quad$ and $\quad \bar{u}_h = \sum_{i \in s_h} \frac{u_{hi}}{n_h}$

The residual $e_{[Y]dk}$ is defined as:

$$e_{[Y]dk} = y_{dk} - \mathbf{x}'_k \hat{\mathbf{B}}_{[Y]d}$$

where: $\quad \hat{\mathbf{B}}_{[Y]d} = \left(\sum_{k \in s} a_k \mathbf{x}_k \mathbf{x}'_k / c_k \right)^{-1} \sum_{k \in s} a_k \mathbf{x}_k y_{dk} / c_k$

and: $\mathbf{x}'_k$ is the transpose of the auxiliary vector $\mathbf{x}_k$, c_k is the model variance of element k, and y_{dk} is the response variable for element k in domain d.

Other variables are defined on page 5 of the WFJ QMI.

Benchmarking

The formula shown in eq.2 is adjusted using an approximation that attempts to account for the benchmarking process. Denoting the associated variance for the current period (c) in a given domain for STES as $\hat{V}_c$, the BRES estimate for that domain at the time of the last benchmark point (b) as $\hat{Y}_{b,BRES}$, and the corresponding (not benchmarked) STES estimate at time b as $\hat{Y}_{b,STES}$, we obtain:

$$\hat{V}_{c,EJ} = \hat{V}_c \times \max \left\{ \left(\frac{\hat{Y}_{b,BRES}}{\hat{Y}_{b,STES}} \right)^2, 1 \right\} \qquad \text{(eq. 3)}$$

where $\hat{V}_{c,EJ}$ is the approximate variance of EJ for a given domain at time c.

This adjustment is based on an approximation that treats the benchmarked estimate as the not benchmarked estimate multiplied by a constant (the ratio of the most recent benchmark to the STES estimate from the same period). The constant is therefore squared in the variance calculation. In reality, the ratio is not a constant because both its components are sample estimates. In principle, the variance of the ratio itself should be estimated and incorporated into the formula. However, if the variance of the

ratio is small then its impact on the formula is negligible and the approximation is reasonable. The max function ensures that the estimated variance for EJ is not lower than the estimate for STES. This is because the ratio is only an approximation, and using it to reduce the variance would result in less conservative confidence intervals.

Sampling errors for BRES are published on the ONS website alongside corresponding employment estimates.

Local unit approximation

It should be noted that the variable y_i is only a proxy for the employment of local unit i. It is calculated by applying auxiliary information from the Inter-Departmental Business Register (IDBR) to the sampled reporting unit return to which local unit i belongs (again, see QMI for further details). Essentially, the variance calculation assumes that y_i is a true return. So far it has not proved possible to estimate the amount of bias introduced by the local unit approximation process.

Imputation

Another assumption made by GES is a 100% survey response rate. In reality approximately 10% of STES is imputed due to non-response, based on movements in similar types of businesses.

Outliers

The outlier treatment used in the production of quarterly EJ estimates is removed prior to the generation of variances so as to avoid providing a misleading level of precision.

Self-employment Jobs component

Self employed jobs are sourced from the Labour Force Survey (LFS). Information on sampling errors for the LFS can be found in section 8 of the Labour Force Survey user guide, volume 1.

For further information on sampling variability for Workforce Jobs please contact Steffan Hess:
☎ 01633 456714 | ✉ steffan.hess@ons.gsi.gov.uk

Confidence intervals - region by industry section

Workforce Jobs

	SIC 2007 Section	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East of England	London	South East	South West	Wales	Scotland	Northern Ireland	United Kingdom
A	Agriculture, forestry & fishing	4	12	10	12	15	13	1	11	14	12	18	-	39
B	Mining & quarrying	1	2	0	0	2	0	2	2	2	2	4	-	6
C	Manufacturing	5	11	22	11	12	11	16	19	14	8	9	-	44
D	Electricity, gas, steam & air conditioning supply	0	2	0	2	0	1	1	2	0	0	2	-	4
E	Water supply, sewerage, waste & remediation activities	1	3	3	2	2	4	3	4	2	2	2	-	8
F	Construction	8	16	12	14	14	16	26	19	18	10	12	-	52
G	Wholesale & retail trade; repair of motor vehicles and motorcycles	10	25	16	16	33	21	30	25	18	13	16	-	69
H	Transport & storage	6	10	7	9	11	11	15	13	15	8	8	-	35
I	Accommodation & food service activities	9	23	10	17	8	20	32	37	82	13	32	-	95
J	Information & communication	3	15	6	9	22	14	83	22	15	4	8	-	90
K	Financial & insurance activities	1	6	6	135	10	7	25	8	6	3	8	-	143
L	Real estate activities	2	10	8	6	5	5	11	10	7	4	8	-	22
M	Professional scientific & technical activities	8	25	12	14	21	20	43	25	15	21	22	-	70
N	Administrative & support service activities	13	18	32	14	12	19	34	22	19	14	17	-	68
O	Public admin & defence; compulsory social security	2	2	5	2	1	3	4	5	3	2	2	-	10
P	Education	11	12	8	10	11	27	16	14	8	24	6	-	46
Q	Human health & social work activities	13	18	15	10	16	22	35	21	17	23	14	-	61
R	Arts, entertainment & recreation	5	10	9	7	45	9	18	13	14	7	8	-	59
S	Other service activities	7	55	10	7	10	17	13	19	8	6	11	-	59
	All jobs	29	74	53	147	73	62	121	75	81	46	56	19	266

Source: Office for National Statistics

Notes

1. It is expected that in 95% of samples, the given range would contain the true value. For example, if the Workforce Jobs estimate for a particular domain is 10,000 and the corresponding figure in the above table is 2,000, then it is expected that in 95% of samples, the true value would lie between 8,000 and 12,000.
2. Workforce Jobs sampling variability is approximated by combining variances for private sector employee jobs from the Short term Employer Surveys (STES) and self employed jobs from the Labour Force Survey (LFS).
3. Approximate sampling errors for Northern Ireland employee jobs are only available at a whole economy level.
4. The outlier treatment used in the production of quarterly Employee Jobs estimates is removed prior to the generation of variances so as to avoid providing a misleading level of precision. In some cases this may lead to extreme values.

Coefficients of variation - region by industry section

Workforce Jobs

<div align="right">Coefficients of variation (%)
Reference period: June 2012</div>

SIC 2007 Section	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East of England	London	South East	South West	Wales	Scotland	Northern Ireland	United Kingdom
A Agriculture, forestry & fishing	28.8	22.6	16.0	12.6	20.5	13.1	48.6	13.3	13.7	21.8	17.4	-	4.8
B Mining & quarrying	18.4	39.4	4.4	0.9	29.5	9.6	21.7	20.5	21.2	32.7	5.2	-	4.4
C Manufacturing	2.2	1.9	4.2	2.1	2.1	2.2	6.5	3.3	2.9	2.8	2.2	-	0.9
D Electricity, gas, steam & air conditioning supply	0.0	8.6	0.9	5.9	0.3	9.2	6.0	5.4	1.0	1.4	4.0	-	1.7
E Water supply, sewerage, waste & remediation activities	11.1	6.5	7.3	8.7	4.2	10.0	7.2	6.8	4.2	5.7	4.9	-	2.1
F Construction	6.0	3.7	4.1	4.8	4.4	3.9	5.0	3.5	5.2	5.6	3.5	-	1.3
G Wholesale & retail trade; repair of motor vehicles and motorcycles	3.0	2.4	2.2	2.2	3.8	2.2	2.4	1.9	2.2	3.4	2.2	-	0.7
H Transport & storage	5.9	3.2	2.9	4.0	4.2	3.9	2.8	3.1	6.9	8.7	3.4	-	1.2
I Accommodation & food service activities	5.8	4.9	3.3	6.3	2.8	5.9	4.5	6.3	22.5	6.5	8.0	-	2.3
J Information & communication	4.9	7.7	4.3	8.5	15.5	6.6	11.8	4.3	8.9	7.2	5.1	-	3.7
K Financial & insurance activities	1.8	3.1	3.6	178.4	7.8	4.7	3.4	3.1	3.3	4.4	4.8	-	6.4
L Real estate activities	6.6	10.6	17.7	11.7	7.5	6.8	6.5	8.6	7.3	14.8	14.1	-	2.6
M Professional scientific & technical activities	6.1	5.2	3.9	6.2	6.7	4.5	3.6	2.9	3.7	16.5	5.1	-	1.4
N Administrative & support service activities	8.2	3.6	8.0	3.8	2.8	3.9	3.6	3.2	5.2	7.8	4.5	-	1.4
O Public admin & defence; compulsory social security	1.6	0.7	1.8	0.9	0.6	1.5	1.0	1.2	1.0	0.9	0.6	-	0.3
P Education	5.2	2.2	1.6	2.7	2.5	5.6	2.3	1.8	1.7	8.3	1.6	-	0.9
Q Human health & social work activities	3.7	2.0	2.3	1.9	2.4	3.5	3.5	2.0	2.4	5.8	1.9	-	0.8
R Arts, entertainment & recreation	7.5	6.3	7.4	5.8	34.7	5.6	5.4	4.7	9.3	10.8	5.1	-	3.4
S Other service activities	13.2	31.3	7.5	6.0	7.8	9.3	5.0	6.4	5.8	8.1	8.4	-	3.4
All jobs	1.3	1.1	1.1	3.4	1.4	1.1	1.2	0.8	1.5	1.7	1.1	1.2	0.4

Source: Office for National Statistics

Key

CV < 5%	
CV => 5% and < 10%	
CV => 10% and < 20%	
CV => 20%	

Notes

1. The colour coding within the tables indicates the quality of each estimate based on the coefficient of variation (CV) of that estimate.

2. The CV is the ratio of the standard error of an estimate to the estimate itself; the smaller the coefficient of variation the greater the accuracy of the estimate. CVs that are greater or equal to 20% should be used with caution.

3. Workforce Jobs sampling variability is approximated by combining variances for private sector employee jobs from the Short term Employer Surveys (STES) and self employed jobs from the Labour Force Survey (LFS).

4. Approximate sampling errors for Northern Ireland employee jobs are only available at a whole economy level.

5. The outlier treatment used in the production of quarterly Employee Jobs estimates is removed prior to the generation of variances so as to avoid providing a misleading level of precision. In some cases this may lead to extreme values.