



Department for
Digital, Culture,
Media & Sport

DCMS Sectors Economic Estimates: Business Demographics Quality Assurance Report

Contents

Introduction.....	2
Quality assurance processes at ONS.....	2
Questionnaire design.....	2
Sampling frame	2
Sample design.....	3
Data collection and response rates.....	3
Editing and validation	3
Imputation	4
Estimation of totals	4
Outliers.....	4
Post-results processing validation	4
Sampling errors	4
Further information	5
Quality assurance processes at DCMS.....	5
Data requirements.....	5
Checking of the data delivery	5
Data analysis.....	5
Quality assurance of data analysis	6
Dissemination.....	6
Next steps.....	6

Introduction

This document summarises the quality assurance processes applied during production of the **DCMS Sectors Economic Estimates 2018: Business Demographics** release (published 29th May 2020). The release covers the number and characteristics of businesses in DCMS sectors, and is based on data provided to DCMS by Office for National Statistics (ONS)¹. This document covers quality assurance carried out by both DCMS and our data providers (ONS).

Quality assurance processes at ONS

The data underpinning this release are taken from the Office for National Statistics (ONS)¹ Annual Business Survey (ABS)^{2,3}. The survey covers the UK non-financial business economy, which accounts for approximately two-thirds of the economy in terms of gross value added (GVA).

Quality assurance at ONS takes place at a number of stages. The various stages and the processes in place to ensure quality at each stage are outlined below. This information is taken from the 'Annual Business Survey technical report: November 2018'⁴ and should be credited to colleagues at ONS.

Questionnaire design

The ABS covers a wide range of diverse industries. To account for this diversity, ONS produces 48 different questionnaire types so that respondents only receive questions applicable to their specific business area. The questionnaires also come with detailed instructions on how to complete (and return) them. This reduces survey burden compared to sending a generic questionnaire that covers all industries, as respondents don't have to sift through a number of questions which to them may be irrelevant. The effect of this is to improve the likelihood that the correct sections/questions of the form are filled and with fewer errors.

Sampling frame

The sampling frame for the ABS is the Inter-departmental Businesses Register (IDBR). Businesses appear in the IDBR if they are registered for Value Added Tax (VAT) with Her Majesty's Revenue and Customs (HMRC), registered for a Pay As You Earn (PAYE) scheme with HMRC, or are an incorporated business registered at Companies House. The IDBR covers businesses in all parts of the economy, except some very small businesses; the self-employed and those without employees, both of which are not registered for PAYE, and those with low turnover, which are not registered for VAT; and some non-profit making organisations. It is kept up to date using supplementary data from surveys such as the Business Register and Employment Survey (BRES).

¹ <https://www.ons.gov.uk/>

² <https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualbusinesssurvey>

³ NB: Whilst ONS are the providers of the data underpinning the Business Demographics releases, they bear no responsibility for their further analysis and interpretation within the releases.

⁴

<https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/methodologies/annualbusinesssurveytechnicalreportaugust2018#sampling-procedure>

Sample design

Data are collected by Office for National Statistics (ONS) from around 62,000 businesses in Great Britain and by the Northern Ireland Statistics and Research Agency (NISRA) from around another 11,000 businesses in Northern Ireland. Sample selection is carried out using a stratified random sample design defined using three strata: employment size band; SIC; and geographical region.

Each sample is generally selected for two years and there is a year-to-year overlap of half the sample. That is, in any year, half of the sample will be newly-selected, and half will have been selected in the previous year as well. This design means that, for half the sample, returns are available from the same businesses in consecutive years and this helps to maintain the quality of editing and validation, imputation and outlier detection.

A sample re-optimisation is carried out every five years to improve the efficiency of the sample estimation and reduce sampling variability as part of the regular process to improve estimates.

Data collection and response rates

ABS questionnaires for a given year are printed for a staggered despatch between January and February of the following year. In order to meet the minimum accuracy standards required by its users, the ABS questionnaire response rate target is at least 64% of businesses by the end of August and 74% by the end of December. If businesses who have received questionnaires have not responded by the deadline, up to three reminder letters can be sent, with the exact approach varying according to the business' sector and/or the importance for data quality of capturing that business (e.g. this may be determined by business size in terms of employees or turnover). There is also an enforcement strategy in place for use as a last resort when all other attempts to encourage a response from a business have failed.

Editing and validation

Once questionnaire responses have been received, they are electronically scanned into the data store. This reduces the risk of error compared to manual human input of the data.

Data are then transferred to the processing system. Initial validation checks are carried out on the returned data. For example, data will fail validation if:

- The data are for periods other than the required year.
- The questionnaire is not the correct type for the business responding.
- There is an invalid question number on the questionnaire.
- No questions have been completed.

After this initial validation further editing is carried out, such as:

- The sums of breakdowns are checked against the totals they contribute to.
- Automatic rounding where actual (non-rounded) values have been provided instead of the required rounded values.
- Date tests to ensure the period covered by the survey return falls into an acceptable range.

Selective editing (using SELEKT Tool) is also carried out. This is an automated process for flagging errors according to:

- Suspicion of an error or mistake.
- Potential impact on estimate.

- Importance of the variable.

Using SELEKT increases the efficiency of the editing process by focusing on the responses with the highest impact and importance.

Imputation

Imputation refers to the estimation of values missing due to non-response. For businesses that have not responded to the ABS, imputation is used as opposed to deletion (where all subjects with any missing values are omitted from the analysis). The method uses returned values from similar businesses to estimate values for non-responding sampled businesses. Imputations are done mainly for large businesses such as those in size band 6 (250 or more employment) and businesses with low employment, but high turnover.

Estimation of totals

The ABS is a survey and, as such, collects information from a sample of the UK business population each year. Results from the survey are scaled up to give information on the UK business population as a whole, using a technique called weighting. A standard weighting method is used for the ABS with additional adjustments to account for businesses that either cease trading (die) or begin trading (are born) in the time between the sample being selected and the questionnaires being sent out.

Outliers

Outliers in the ABS data are identified using both an automated system and manually (by expert judgement by ONS staff). These outliers are then treated with a method called post-stratification, in which weighting for the outliers is reduced so that they do not have a large effect on the overall estimates. Weights for the remaining businesses in the stratum are recalculated once the outlier has been removed.

Post-results processing validation

Post-results processing validation refers to the stage where checks are done on the final industry results before publishing. These checks are carried out by different teams in charge of the specific industries; however, for all the teams the checks are done in a similar way. These checks include comparison of trends in the current year's data compared to the previous year's data. If there is an atypical result, this will be followed up to find out if there were any specific one-off events that might have caused this. The atypical result is treated according to any further information found and the results for other businesses in the same stratum.

Sampling errors

In the interest of transparency, ONS publish sampling errors down to the 4 digit Standard Industrial Classification (SIC) code level⁵. These give an indication of the level of uncertainty around the figures.

⁵ The sampling errors are available at:

<https://www.ons.gov.uk/file?uri=/businessindustryandtrade/business/businessservices/datasets/uknonfinancialbusinesseconomyannualbusinesssurveyqualitymeasures/current/absqualitymeasures.xls>

Further information

Further technical information about the ABS, including how the quality of the data is assured is available in the technical report⁶.

Quality assurance processes at DCMS

The majority of quality assurance of the data underpinning the DCMS Sectors Economic Estimates Business Demographics release takes place at ONS, through the processes described above. Once ONS have ensured all their in-house data checks, the data required by DCMS are sent via secure transfer. Further quality assurance checks are then carried out within DCMS.

Production of the analysis and report is typically carried out by one member of staff, whilst quality assurance is completed by at least one other, to ensure an independent evaluation of the work.

Data requirements

DCMS discussed our data requirements with ONS and these are formalised as a Data Access Agreement (DAA). The DAA covers which data are required, the purpose of the data, and the conditions under which ONS provide the data. Discussions of requirements and purpose with ONS improved the understanding of the data at DCMS, helping us to ensure we receive the correct data and use it appropriately.

Checking of the data delivery

DCMS check that the data delivered by ONS match what is listed in the Data Access Agreement (DAA). For this particular release we check that:

- We have received all data at the 4 digit SIC code level, which is required for us to aggregate up to produce estimates for our sectors.
- Data at the 4 digit SIC code has not been rounded unexpectedly. This would cause rounding errors when aggregating up to produce estimates for our sectors.
- Where bespoke breakdowns are required (the employment and turnover sizebands used by DCMS differ to those used in the ONS publication) that these have been defined correctly.

Data analysis

At the analysis stage, data are aggregated up to produce information about DCMS sectors. The Business Demographics lead builds in the following checks at this stage:

- Checks that summing up breakdowns gives the same figure as the total they contribute to. E.g.:
 - If you sum together businesses in all employment bands in each sector do you get the total number of businesses in that sector?
 - Does the number of foreign-owned businesses equal the number of EU (excluding UK)-owned businesses plus the number of non-EU-owned businesses?
- “Sense checks” of the data. E.g.:

⁶

- Does the number of businesses in DCMS sectors look similar to last year? If not, could this be because of changes to the number of businesses in the UK non-financial business economy as a whole (like rapid growth/decline) or changes to the ABS methodology?
- Do the businesses in DCMS sectors have the same/similar distribution in terms of size and geographical location as previous years?

Quality assurance of data analysis

Once analysis is complete, the producer hands over to the quality assurers to carry out further checks of the analytical work completed. Suggested checks are recommended by the producer in the form of a document called a quality assurance (QA) log. Within the document, the suggested checks are listed and described, and there is space for the quality assurers to indicate the outcome of the checks and give any other feedback. After the publication, the quality assurance processes are reviewed to ensure they are relevant and comprehensive. The checks listed in the QA log cover:

- Ensuring the correct data have been used for the analysis e.g. has the 2016 data been used to derive the 2016 figures for DCMS sectors, or has the 2017 data been used by mistake?
- Checking that the correct SIC codes have been aggregated together to form DCMS sector estimates. Are all SIC codes we require included? Are there any non-DCMS SIC codes that have been included by accident?
- Making sure it is not possible to derive disclosive data from the figures that will be published.

Dissemination

Finalised figures are disseminated within Excel tables and a written report (which includes written text, graphs, tables and infographics) published on GOV.UK. These are produced by the Business Demographics lead. Before publishing, a quality assurer checks the figures match between the working-level analysis, the tables and the written report. The quality assurer also makes sure any statements made about the figures (e.g. regarding trends) are correct according to the analysis.

Next steps

We encourage our users to engage with us so that we can improve our statistics and the documentation surrounding them. If you would like to comment on this first report on quality assurance processes, or have any enquiries please get in touch at evidence@culture.gov.uk.



Department for
Digital, Culture,
Media & Sport

4th Floor

100 Parliament Street
London
SW1A 2BQ



© Crown copyright 2019

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/ or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email:

psi@nationalarchives.gsi.gov.uk