Department for Digital, Culture, Media & Sport

DCMS Sectors Economic Estimates: Regional GVA Quality assurance report

Contents

Introduction

The **DCMS Sectors Economic Estimates 2017 (provisional): Regional GVA** statistics were published on 13th February 2019 and provides an estimate for the contribution of DCMS sectors to each region in the UK, represented by Gross Value Added (GVA). This document summarises the quality assurance processes applied during the production of these statistics and includes a detailed account of the quality assurance processes and the data checks carried out by our data providers (Office for National Statistics)[1] as well as by DCMS.

Data Sources

The following data sources were used in the production of Regional GVA estimates for DCMS sectors:

1. ONS Regional Gross Value Added balanced tables (published 12 December 2018) – containing data up to 2017[2].
2. Unsuppressed Annual Business Survey (ABS) approximate Gross Value Added (aGVA) estimates at the lowest level available – used to allocate GVA below the level available in the supply and use tables[3].

Regional GVA estimates are consistent with the Office for National Statistics balanced regional gross value added (i.e. GVA(B)) estimates. As the DCMS sectors are defined at 4 digit SIC level, a method for breaking down the regional estimates of GVA was applied. This is calculated by using aGVA data from the UK non-financial business economy ABS survey.

Quality assurance processes at ONS

Quality assurance at ONS takes place at a number of stages. The various stages and the processes in place to ensure quality for both data sources are outlined below. It is worth noting that information presented here on the data sources are taken from the Regional GVA technical report[4] and the Annual Business Survey technical report[5] and should be credited to colleagues at the ONS.

ONS Regional Gross Value Added balanced tables

Regional Gross Value Added is the value generated by any unit engaged in the production of goods and services. It is measured at current basic prices, excluding taxes (less subsidies) on products. GVA plus taxes (less subsidies) on products is equivalent to Gross Domestic Product (GDP).

ONS produce these estimates of regional gross value added using estimates from gross value added income (GVA(I)) and gross value added production (GVA(P)) to produce a balanced measure of regional GVA, known as GVA(B). This gives users a single measure of economic activity within a region and are compiled on a workplace basis (allocated to the location where the economic activity takes place).

1. **The income approach (GVA(I))** involves adding up the income generated by resident individuals or corporations in the production of goods and services. It is calculated gross of deductions for consumption of fixed capital, which is the amount of fixed assets used up in the process of production in any period.

---

[1] https://www.ons.gov.uk/

[2] https://www.ons.gov.uk/economy/grossvalueaddedgva/bulletins/regionalgrossvalueaddedbalanceduk/1998to2017

[3] https://www.ons.gov.uk/businessindustryandtrade/business/businessservices

[4] https://www.ons.gov.uk/economy/grossvalueaddedgva/bulletins/regionalgrossvalueaddedbalanceduk/1998to2017#quality-and-methodology

[5]https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/methodologies/annualbusinesssurveytechnicalreportaugust2018#sampling-procedure

2. **The Production approach (GVA(P))** is measured at both current prices and in chained volume measures (CVM). It is calculated for a given reference period as the total value of all goods and services produced (output), less goods and services used up or transformed in the production process, such as raw materials and other inputs (intermediate consumption). The production approach to compile GVA is conceptually equivalent to the income approach, but allows deflation of current prices to produce constant price measures, since the production components relate to goods and services that can be broken down into price and volume indices.

Main data sources for GVA (Income):

The primary input datasets for the compilation of regional GVA(I) for all UK sectors are listed below. They include administrative data and data from structural surveys. The data are acquired from both internal (ONS) and external sources.

Table 1:Summary of main data sources for indicators used in Regional apportionment

Data source	Nature of data	Use in Regional Accounts
Annual Survey of Hours and Earnings (ASHE)	An ONS workplace-based sample survey of wages and salaries	Allocates compensation of employees (CoE)
Business Register and Employment Survey (BRES)	An ONS structural business survey, a sample survey of enterprises	Allocates CoE
Annual Business Survey (ABS)	An ONS structural business survey, a sample survey of enterprises	Allocates mixed income, CoE and gross trading profits of other corporations
Agricultural Accounts from Department for Environment, Food and Rural Affairs (Defra)	Structural business survey of the agricultural sector, a sample survey	Used to allocate rental income, CoE, mixed income, gross trading profits of partnerships, and holding gains, all in the agricultural sector
Labour Force Survey (LFS)	An ONS residence-based sample survey answered by employees	Used to apportion CoE data for the activities of households industry
Defence Analytical Systems and Advice (DASA)	Administrative data about the number and location of armed forces personnel	Used to apportion national estimates of forces pay by region and Extra-Regio
BEIS	Administrative data about onshore gas and oil profits	Used to apportion gross operating surplus

Self-Assessment income tax data from HM Revenue and Customs (HMRC)	Administrative data provided by HMRC, the national tax authority. Derived from self-assessment tax returns completed by self-employed sole traders and partners.	Used to apportion gross trading profits of partnerships, and mixed income

ONS currently publishes estimates of regional GVA(I) at NUTS1, NUTS2 and NUTS3 levels annually in December. The UK Regional GVA estimates are constrained to the latest published UK National Accounts Blue Book. The Blue Book is the annual publication of UK National Accounts by ONS.

Data sources used as regional indicators are collated, analysed and validated. Outliers are identified using graphical analysis and quality adjustments are assigned where necessary. Where appropriate, issues are referred back to the suppliers. The resulting validated datasets are then subject to a peer review process so that their impact can be assessed, prior to their inclusion in the compilation of GVA(I).

Main data sources for GVA (Production):
The primary input datasets for the compilation of regional GVA(P) are listed below in Table 2. The data are acquired from both internal (ONS) and external sources.

Table 2: Summary of main data sources for indicators used in Regional apportionment

Data source	Nature of data	Use in Regional Accounts
Annual Business Survey (ABS)	An ONS structural business survey, a sample survey of enterprises	Allocates output and intermediate consumption for bulk of the UK economy
Annual Survey of Hours and Earnings (ASHE)	An ONS workplace-based sample survey of wages and salaries	Used in conjunction with BRES data to allocate public expenditure
Business Register and Employment Survey (BRES)	An ONS structural business survey, a sample survey of enterprises. Used to define regional public employment numbers	Used in conjunction with ASHE data to allocate public expenditure
Agricultural Accounts from Department for Environment, Food and Rural Affairs (Defra)	Structural business survey of the agricultural sector, a sample survey	Used for output and intermediate consumption for part of agricultural sector (crops and livestock)
Self-Assessment (Pay As You Earn and Self Assessment), Her Majesty's Revenue and Customs (HMRC)	Administrative data provided by HMRC, the national tax authority. Derived from self-assessment tax returns completed by self-employed sole traders	This forms an indicator that is used to apportion gross value added for sole traders

Regional estimates of bank and building society fees and commission and FISIM (Bank of England)	Administrative data supplied by the Bank of England at NUTS1	Allocates GVA for monetary intermediation (SIC 64.1)
Labour Force Survey (LFS)	Labour Force Survey data on regional employee numbers in component finance industries	Used as an indicator for remainder of financial services (SIC 64), and temporarily for insurance industries (SIC 65) and auxiliary financial activities (SIC 66)

ONS publish estimates of regional GVA(P) at NUTS1 and NUTS2 levels annually in December. All estimates are constrained to the latest published UK National Accounts Blue Book

Validation and quality assurance for GVA(I)

All regional GVA estimates are constrained to the Blue Book totals. Below the UK level, NUTS1 to NUTS3 estimates are constrained to sum to the UK totals. National totals for GVA components (consistent with the latest Blue Book publication) are collated and allocated to regions using the most appropriate regional indicators. Consistency checks are built into the processing system to ensure that regional aggregates tally with national totals.

The main regional indicators are acquired from published administrative data and structural business surveys (as outlined above). These data are subject to a rigorous validation process that entails a peer review process and a mechanism for referral back to data suppliers where appropriate.

Sample sizes can vary from 100% (HMRC Self assessment data) to 1% (HMRC PAYE data). These variations are unavoidable in the collation of reliable time series data for the calculation of GVA. Improvements in the GVA methodology or changes in administrative source data allows for reviewing the available sources for higher quality datasets or more timely publications. This is an ad hoc and continuous process which does not include major revisions to the methodology of GVA.

The vast majority of source data are annually updated, however, some datasets are published biennially or on an ad hoc basis. These missing values are imputed, using other available information. As with sample size, potential improvements to the methodology are reviewed whenever new data become available or when significant changes to the source data affect the final values. Where no recent estimate is available, the previous year's data may be used. It is important to note however, that these issues are rare because of the completeness of the main source data.

Validation and quality assurance for GVA(P)

As with GVA(I), checks are built into the processing system to ensure that regional aggregates tally with national totals.

The important regional indicators are acquired from published administrative data and structural business surveys. These data are subject to a rigorous validation process that entails a mechanism for referral back to data suppliers where appropriate.

Sample sizes can vary. These variations are unavoidable in the collation of reliable time series data for the calculation of GVA(P). Improvements in the GVA(P) methodology or

changes in administrative source data allow for reviewing the available sources for higher quality datasets or more timely publications. This is an ad hoc and continuous process that does not include major revisions to the methodology of GVA(P).

The vast majority of source data are updated annually, but some datasets are published biennially or on an ad hoc basis. These missing values are imputed as stated in the validation process. As with sample size, potential improvements to the methodology are reviewed whenever new data become available or when significant changes to the source data affect the final values. Where no recent estimate is available, the previous year's data may be used. It is important to note, however, that these issues are rare because of the completeness of the main source data.

Input datasets are compared to the data received in previous years' deliveries. Large fluctuations against previous years are investigated. Following investigation, any anomalous data that remain unexplained are followed up, where appropriate, with the relevant data providers for clarification.

The regional accounts team then makes adjustments to the source datasets where necessary. The output variables for publication are subject to rigorous scrutiny, including looking at growth and shares, graphical depictions and comparisons with previous data. Resulting anomalies are investigated and rectified where necessary. Feedback from this process is documented on the regional accounts database and any actions implemented within the results.

Revisions for GVA(I) and GVA(P):

The complex process by which the GVA estimates are produced means that it is not currently possible to define the accuracy of the estimates in terms of detailed statistical properties, for example through their standard errors. Therefore, the reliability of the estimates is measured by the extent of revisions. Headline UK estimates will be affected by revisions to the UK National Accounts Blue Book, while revisions to NUTS1 and lower level estimates will come from revisions to the data used to apportion the UK GVA(I) estimates and revisions to the Blue Book.

Estimates of approximate GVA are also published as part of the Annual Business Survey release. These estimates are used in the production of annual Supply and Use Tables for the compilation of the UK gross domestic product (GDP). The ABS GVA estimates differ from the UK Regional Accounts estimates in a number of ways:

- Regional Accounts GVA(P) estimates are constrained to National Accounts Blue Book totals

- Regional Accounts GVA(P) estimates are analysed and produced as a time series

- ABS estimates represent the UK business economy, which is around two-thirds of the UK economy but excludes agriculture, public sector and financial sector.

As the DCMS sectors are defined at 4 digit SIC level, a method for breaking down the regional estimates of GVA was applied. This is calculated by using aGVA data from the UK non-financial business economy ABS survey. The next section outlines the ABS quality assurance process.

Annual Business Survey

As the DCMS sectors are defined at 4 digit SIC level, and the regional GVA table is at a higher level, a method for breaking down the regional estimates of GVA was applied. This was done using approximate Gross Value Added (aGVA) data from the UK non-financial business economy (Annual Business Survey).

The Annual Business Survey (ABS) measures business and financial information from UK businesses, including total turnover, total employment costs, total purchases, capital expenditure, stocks and other aggregates. Variables derived from these statistics, such as approximate gross value added at basic prices are also published by the ABS.

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).

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 re-calculated 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 level[6]. These give an indication of the level of uncertainty around the figures.

Approximate gross value added (aGVA) vs GVA in the National Accounts

aGVA represents the amount that individual businesses, industries or sectors contribute to the economy. Generally, this is measured by the income generated by the business, industry or sector minus their intermediate consumption of goods and services used up in order to produce their output, labour costs (for example, wages and salaries) and an operating surplus (or loss). The latter is a good approximation for profits, from which the cost of capital investment, financial charges and the payment of dividends to shareholders are met.

There are differences between the approximate measure of aGVA calculated by ABS and the measure of gross value added (GVA) used in the national accounts. The National Accounts carry out coverage adjustments, conceptual adjustments and coherence adjustments. This estimate of GVA uses inputs from a number of surveys and covers the whole UK economy, whereas the ABS does not include some parts of the agriculture and financial activities sectors, or public administration and defence.

The ABS total aGVA is around two-thirds of the national accounts whole economy GVA because of these differences. Real (inflation-adjusted) estimates of national and regional GVA are published in the national accounts and regional accounts respectively. However, national and regional estimates of aGVA from the ABS are not adjusted for inflation.

There are also some conceptual differences between the two measures of gross value added. For example, some production activities such as illegal smuggling of goods must be included in the national accounts but are outside the scope of the ABS.

More information on this can be found in the ABS technical report.

---

[6] The sampling errors are available at:
https://www.ons.gov.uk/file?uri=/businessindustryandtrade/business/businessservices/datasets/uknonfinancialbusinesseconomyannualbusinesssurveyqualitymeasures/current/absqualitymeasures.xls

Quality assurance processes at DCMS

The majority of quality assurance of the data underpinning the DCMS Sectors Economic Estimates Regional GVA release takes place at ONS, through the processes described above. However, further quality assurance checks are carried out within DCMS at various stages.

Production of the 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

For the ABS data, 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 and sub-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 and sub-sectors.

Data analysis

At the analysis stage, data are aggregated up to produce information about DCMS sectors and sub-sectors. The Regional GVA statistics 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.:
    - Do sub-sectors within the Creative Industries sum to the Creative Industries total?
    - Does the individual regional figures sum up to the total for that sector?
- "Sense checks" of the data. E.g.:
    - Are the proportions of GVA for each region similar to last year? If not, could this because of changes to the ABS methodology?
    - Looking at any large differences between the data and possible causes to these.

Quality assurance of data analysis

Once analysis is complete, DCMS document the checks needed for quality assurers to carry out. The checks cover:

- Ensuring the correct data are used for the analysis. For example:
    - Checking the updated UK National Accounts data been used, including the revised back series data.
    - Checking that the sum of the individual regions sum to the national total.
    - Sense checking whether the data looks sensible i.e. comparing against previous year's figures.

- • Checking that the correct SIC codes have been aggregated together to form DCMS sector (and sub-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.
- • Making sure the correct data has been pasted to the final tables for publication and are formatted correctly.
- • Making sure all charts are linking to the right data and all maps produced are using the correct data.

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 Regional GVA statistics 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 and checks for spelling or grammar errors.

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 quality assurance report, 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

OGL