

Quality assurance of administrative data (QAAD) report for regional gross value added (GVA): October 2017

An assessment of the administrative data source used in the production of regional gross value added (GVA).

Contact:
Trevor Fenton
trevor.fenton@ons.gsi.gov.uk
+44 (0)1633 456083

Release date:
30 October 2017

Next release:
To be announced

Table of contents

1. [Background to the statistics](#)
2. [Quality assurance of administrative data \(QAAD\) assessment](#)
3. [Areas of quality assurance of administrative data \(QAAD\)](#)
4. [Summary](#)

1 . Background to the statistics

The Regional Accounts team in the National Accounts and Economic Statistics (NAES) group within the Office for National Statistics (ONS) uses data supplied by Her Majesty's Revenue and Customs (HMRC) in the regional distribution of gross value added (income approach) (GVA(I)) and gross value added (production approach) (GVA (P)).

For regional GVA(I), HMRC self-assessment (SA) data for partnerships and sole traders are used respectively as regional indicators for the gross trading profits (GTP) of partnerships and mixed income (MI) of sole traders components.

For regional GVA(P), HMRC self-assessment (SA) data for sole traders are used as a regional indicator for the GVA of sole traders, for a subset of industries identified as having a significant proportion of self-employed workers.

This report outlines the process taken from initial collection through to the output of the release. It identifies potential risks in data quality and accuracy as well as details of how those risks are mitigated. This report produced by NAES investigates the administrative data source we use in the production of regional GVA as set out by the [UK Statistics Authority](#). As such, this report specifically focuses on our administrative data use of the HMRC self-assessment data.

Further information relating to quality and methodology for the regional GVA can be found in our regional quality and methodology information (QMI) documentation for [GVA\(I\)](#) and [GVA\(P\)](#).

2 . Quality assurance of administrative data (QAAD) assessment

UK Statistics Authority QAAD toolkit

The assessment of our administrative data source has been carried out in accordance with the UK Statistics Authority Quality Assurance of Administrative Data (QAAD) Toolkit.

The administrative data source investigated has been evaluated according to the toolkit's risk and profile matrix (Table 1) reflecting the level of risk to data quality and the public interest profile of the statistics.

Table 1: UK Statistics Authority quality assurance of administrative data (QAAD) risk and profile matrix

Level of risk of quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of a low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
High	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

Source: Office for National Statistics

The toolkit outlines four specific areas for assurance and the rest of this report will focus on each of these areas in turn. These are:

- operational context and administrative data collection
- communication with data supply partners
- quality assurance principles, standards and checks applied by data suppliers
- producer’s quality assurance investigations and documentation

2.2 Assessment and justification against the QAAD risk and profile matrix

Table 2: QAAD risk and profile matrix assessment of administrative self-assessment data

	Low A1	Medium A2	High A3
Operational Context and Administrative Data Collection		[A2]	
Communication with Data Supply Partners		[A2]	
Quality Assurance Principles, Standards and Checks by Data Supplier		[A2]	
Producers Quality Assurance Investigations and Documentation		[A2]/[A3]	

Source: Office for National Statistics

The risk of quality concern and public interest profile has been set as “medium” due to the data in regional gross value added (income approach) (GVA(I)) and gross value added (production approach) (GVA(P)) having wide use and economic interest, but they are not considered market or politically sensitive.

3 . Areas of quality assurance of administrative data (QAAD)

3.1 Operational context and administrative data collection (QAAD score A2)

This relates to the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being compiled and the factors that might increase the risks to the quality of the administrative data.

HM Revenue and Customs (HMRC) is a non-ministerial department of the UK government whose responsibilities include the collection of taxes, self-assessment data, National Insurance contributions, payment of state support and enforcement of the National Minimum Wage. They report to Parliament through the Treasury minister whose role is to oversee spending.

Self-assessment data for mixed income and gross trading profits

The Regional Accounts team uses self-assessment data from HMRC within the “mixed income” component where it is allocated to regions using self-employment data (profits of sole traders) from HMRC. Mixed income represents income generated by sole traders (self-employed people not registered as partners). In national accounts, their income is considered a mixture of profits and self-paid wages (hence “mixed” income) returned to the business. The data received from HMRC are self-assessment data by industry group and by Nomenclature of Units for Territorial Statistics (NUTS) regions at levels 1, 2 and 3.

In addition to using the self-assessment data within the “mixed income” component of gross value added (income approach) (GVA(I)), the Regional Accounts team also uses HMRC sole traders data in the production of the regional distribution of output for the self-employed in gross value added (production approach) (GVA(P)). This is used for industries that are known to have a high proportion of self-employed workers, who are not covered well by the principal survey data source used to measure output.

The self-assessment data on partnerships are used as a component in the measurement of gross trading profits (GTP) of private business enterprises, another published component of GVA(I). The data on partnerships are combined with data on the profits of other private corporations, which are generally of much greater magnitude than the partnerships component.

HMRC collects information on the profits of sole traders and partnerships via its collection of self-assessment data and provides this in time for our annual December publication. Typically HMRC delivers provisional data for the latest year and revised data for the previous year. Both deliveries are received on a financial year basis and require converting to calendar years within the regional GVA processing system. Once converted the data are lagged by a year compared to the published GVA data. The data are based on an extract of almost 100% of the self-assessment data.

The data provided by HMRC are allocated to regions of the UK according to the usual residence of the person completing the self-assessment form. In regional GVA we allocate GVA according to the place where the activity takes place, what we term a “workplace basis”. While it is likely that a good proportion of self-employed people carry out the bulk of their work within their NUTS1 region of usual residence, this assumption loses credibility as we look at smaller geographic areas. In some industries, such as construction for example, it is known that self-employed tradespeople often travel extensively to the sites where they work.

Strengths

- Available in time for the December regional GVA publication.
- Almost full coverage of the population used in the data.

Weaknesses

- Data delivered are based on financial years and require a calendar year conversion by ONS.
- Provisional data are provided for the latest year so there will inevitably be some revisions the following year.
- Data are provided on a residence basis but are used as an indicator for a workplace-based allocation of activity.

3.2 Communication with data supply partners (QAAD score A2)

This relates to the need to maintain effective relationships with suppliers (through written agreements such as service level agreements or memoranda of understanding). This includes change management processes and the consideration of statistical needs when changes are being made to relevant administrative systems.

The ONS Regional Accounts team is in regular contact, via email, telephone and meetings, with the HMRC Knowledge, Analysis and Intelligence (KAI) team and has a formal agreement in place for the data deliveries.

Formal meetings take place annually (or sometimes more frequently) to allow both parties to review the current agreement, discuss the detail of data being delivered, the methodology used to compile the data and the quality assurance carried out by HMRC KAI.

Data delivery dates are discussed at meetings and agreed via email correspondence. HMRC and Regional Accounts maintain more frequent contact around the time of delivery to discuss timeliness and any potential delays and the impact of these. HMRC provides a covering letter with each dataset detailing the quality assurance measures that have been applied to the data.

The Regional Accounts team is confident that this working relationship with HMRC is sufficient and there is adequate communication during the year with more frequent contact nearer the delivery deadlines.

Strengths

- Regular contact between ONS Regional Accounts team and HMRC KAI team.
- Formal agreement in place with delivery dates agreed in advance.
- Briefing letter received accompanying data delivery.

3.3 Quality assurance principles, standards and checks by data supplier (QAAD score A2)

This relates to the validation checks and procedures undertaken by the data supplier, any process of audit of the operational system and any steps taken to determine the accuracy of the administrative data.

Only taxpayers with a valid postcode are used in the data. HMRC conducts checks for missing or invalid postcodes and corrections are made in the source file if necessary.

Details of the quality assurance conducted on the data at source by HMRC are summarized in this section.

The HMRC KAI team conducts quality assurance on the data, providing covering correspondence with each data delivery describing:

- what data are being delivered (including NUTS levels and industry breakdown)
- the classifications used in the data
- what quality assurance measures have been applied
- any assumptions that have been made
- how the data have been produced
- statistical tests applied to identify disclosive data
- investigation of largest year-on-year changes
- investigation of the highest profits and lowest losses
- removal or adjustment of suspect data
- investigations leading to corrections within the data resulting from quality assurance

Strengths

- Extensive data checks carried out by HMRC.
- Accompanying briefing note detailing these checks.
- Data resupplied if errors are found.

3.4 Producers quality assurance investigations and documentation (QAAD score A2)

This relates to the quality assurance conducted by the statistical producer, including corroboration against other data sources.

In addition to the quality checks applied by HMRC, the ONS Regional Accounts team conducts further quality assurance on the data received from HMRC before using it in the production of outputs. This takes the form of time series and graphical analysis using Excel and comparing year-on-year movements and longer-term trends. If any large growth, contraction, revision, or unusual movement in a particular region and industry is identified, the Regional Accounts team makes use of the following options:

- sending a data query to HMRC KAI team
- using statistical, economic and regional knowledge to inform a quality adjustment to the data
- allowing the data to remain as delivered following investigations

The HMRC self-assessment (SA) data delivered includes a number of suppressed cells in line with HMRC's statistical disclosure policy. To use the data, Regional Accounts require all suppressed cells to have data and impute data for these missing cells based on allocating the industry and region totals and are informed by the previous year's data. This means that some of the data have been estimated within the processing by the Regional Accounts team. It is important to note that, even though these missing data have been imputed as part of the process for compiling regional GVA estimates, no confidential data are published by ONS Regional Accounts.

Once Regional Accounts has completed the quality assurance of the HMRC SA data, they are used in the production system to produce provisional outputs of regional GVA(I) and regional GVA(P). These outputs are then subject to analysis and further quality assurance. If this process generates questions or issues, further adjustments may be made to the HMRC SA data or queries raised with HMRC KAI (as discussed in the previous paragraph) and this process repeated.

Regional Accounts has a quality assurance plan for both regional GVA(I) and regional GVA(P), which are updated throughout the production rounds. These plans are consistent with those used throughout national accounts and allow production teams to log each quality assurance step carried out, whether it is completed to timetable and the reasons for any issues or delays.

Regional Accounts keeps a full audit trail of any previous versions of the data stored on an internal drive. Therefore, if there are any revisions, Regional Accounts is able to identify where and to what extent the change has taken place.

Within GVA(I), the sole traders data accounts for approximately 4% of GVA and partnerships data accounts for approximately 3% of GVA. Within GVA(P), the sole traders data accounts for approximately 3% of GVA (less because fewer industries are covered), so it is an important data source for the regional GVA publications.

Strengths

- Sense checks comparing the data with previous trends.
- Audit trail – ONS keeping historic files.
- All quality assurance overseen by national accounts quality plans.

Weaknesses

- Imputation of missing data following HMRC disclosure control means an inevitable reduction in the accuracy of estimates.

4 . Summary

In investigating the administrative source for self-assessment data, Regional Accounts considers the main strengths of the data for our purpose to be:

- high level of coverage (almost 100% of the population covered by the data)
- high level of quality assurance carried out at source by the HM Revenue and Customs (HMRC) Knowledge, Analysis and Intelligence (KAI) team
- additional quality checks carried out by the Office for National Statistics (ONS) Regional Accounts team

We believe the current limitations of this dataset are:

- the use of data allocated on a residence basis as an indicator of workplace-based activity results in a loss of accuracy in regional gross value added (GVA) estimates, particularly for smaller geographies and in particular highly mobile industries
- the need for imputation of missing data by Regional Accounts, caused by suppression for disclosure control, results in a further loss of accuracy in regional GVA estimates

In constantly seeking to improve our data sources, investigations have taken place looking at other possible sources of data on the self-employed and these are explained in this section.

The publication of [Trends in self-employment in the UK](#) appears to be a one-off article published on 13 July 2016. There are no future articles scheduled on the ONS website and therefore, although it provides a useful insight into self-employment, the data included cannot be considered reliable enough to be used as a comparison over multiple years. There are a number of sources cited for this article – ONS, Labour Force Survey, cross sectional datasets and the authors' own calculations, which are not explained within the article. It is therefore difficult to know how these data have been compiled, which would be necessary to allow meaningful comparisons to be made.

Self-employment data from the HMRC Survey of Personal Incomes (SPI) are not currently used in the gross disposable household income (GDHI) publication, which is generally published in May each year, or the GVA publication, which is published in December each year. Following a review of existing procedures, Regional Accounts found the main reasons for this are:

- the data are typically delivered in the March following the GVA publication in December and therefore would not be timely enough for the GVA publication
- the data from the SPI and self-assessment data (currently used by Regional Accounts), are based on the same extract of the data, however, the SPI data uses a sample of this extract whereas the self-assessment data uses almost 100% of the extract; therefore the self-assessment data delivered for the GVA publication in December contains the latest source data and highest content

After reviewing these other sources of self-employment statistics, and the limitations mentioned previously, we consider the HMRC self-assessment data to be a data source that is largely fit for purpose and fulfils the requirements of an A2 assurance rating.