Office for National Statistics

National Accounts Coordination Division's assessment of quality assurance of administrative data used to compile the quarterly and annual UK National Accounts

Quality assurance report produced by NAES to investigate the administrative data sources used in the production of quarterly and national accounts as set out by the UK Statistics Authority.

Contact:
Robert Kent-Smith
robert.kent-smith@ons.gsi.gov.uk
+44(0)1633 651618

Release date:
27 July 2017

Next release:
To be announced

Table of contents

1 . What are administrative data?

"Administrative data" refers to information collected primarily for administrative reasons (not research). This type of data is collected by government departments and other organisations for registration, transactions and record-keeping, usually when delivering a service. Often, administrators use administrative data for operational purposes – their statistical use is secondary. Consequently, administrative data contain information that is not primarily collected for statistical purposes, but statistical offices make use of. Some examples of administrative data used for statistical purposes come from the following:

- Value Added Tax (VAT) data

- personal Income Tax data

- business (including corporate) taxation data

- social security data

- business registration and administration records

- business accounts of corporations

- records held by the Bank of England

- records (other than VAT) held by HM Revenue and Customs

- records of government (central and local)

- records held by associations of employers, of employees and of businesses and professions

- records held by other private sector bodies, for example, credit-rating agencies, non-profit units

The use of administrative data offers several advantages:

- are "cheaper" than other sources and often even free

- provide complete, or almost complete, coverage of the population to which the administrative process applies often providing more accurate and detailed estimates of sub-populations

- timeliness of the statistical variables derived from administrative sources is improved

- reduce the response burden on survey participants

- may improve business register quality

2 . Use of both public and private (often business) data in the national accounts

Since their inception, national accounts have used a mix of public and private data to provide a comprehensive picture of overall economic activity that is timely and accurate. The UK makes extensive use of partial data, public and private, as extrapolators for its early estimates, sometimes aggregations of business and government microdata. Microdata are used in matching of statistical and non-statistical data to improve official statistics by developing bias adjustments for survey data, improving coverage and by identifying reporting and other problems. For a number of reasons, this pattern of using big data collected for non-statistical purposes, as extrapolators and methodological research and improvement tools, is likely to continue.

3 . Assuring the quality of both business and administrative data

Because business and administrative data are collected for non-statistical purposes, they often do not meet statistical standards in terms of representativeness, concepts, definition, collection methods and so on. The use for statistical purposes of administrative sources requires a careful evaluation of their conceptual base, classification and time reference. So, to use these administrative and business data, national accountants investigate and understand the statistical characteristics of the data and improve the accuracy of these non-statistical extrapolators through weighting, filling in gaps in coverage, bias adjustments, averaging with other extrapolators, and benchmarking and balancing. In evaluating business and administrative data for use in national accounts, the statisticians have to address some important questions.

One of the first questions a national accountant would ask is how closely does the data fit with national accounts concepts? For example, the treatment of tax in turnover definitions can vary between sources. The first step is to understand the administrative data better by talking to the source data holder. In some cases, the differences may be small enough to be acceptable. If bigger differences are understood fully, it may be possible to correct them either directly or through communicating with the source.

A second question that must be addressed in the use of these big data is the consistency of the time frame in the source data with the time frame for the national accounts estimate.

The third issue relates to the representativeness of the external source data and any selection biases that may be present in the data.

4 . Other problems in using data from administrative sources

Although there are many good reasons for using administrative sources, there are also a number of problems.

Gaps in data

Unfortunately, some of the largest gaps in coverage for the official economic statistics are difficult to fill using publicly available data. Significant gaps in official output statistics such as those in services and local government are hard to fill because they are in sectors dominated by a large number of relatively small units using an assortment of concepts, definitions, reporting periods and accountancy rules.

Small firms do not file public reports or make available anything other than their industry, services offered and location (items normally found in business directories), or sales advertised on the internet. Because they are among the items businesses find most sensitive, small firms generally do not post or file information on their sales, prices and costs. Filling gaps in local government data is also difficult. Similarly, gaps in income statistics are in hard-to-fill areas like small business income or self-employed incomes, which are often only reported on individual tax returns.

To construct statistics from administrative data, national accountants usually need to integrate them with statistical data in some way. Even when admin data cover the required population completely, statistical data may be needed to ensure the data can be aggregated to the correct output domains. Matching can often be done via the business register. Integration is not always straightforward and national accountants discuss issues with linking, coverage and definitions.

Linking administrative data to statistical data

Due to historical use of admin data for the business register, it is common to find a unique identifier between admin and survey sources. However, differences in reporting structures often mean that there is not a one-to-one match between sources.

Cleaning administrative data

It is often taken for granted that administrative data are free of errors as they are collected for important official reasons. In practice, errors exist and need to be identified and treated in some way. Traditional survey data editing methods are not always appropriate for administrative data. National accountants cannot re-contact businesses.

Types of error statisticians can find in administrative data

Statisticians can often find two principal types of errors: systematic errors and random errors. Usually, the causes of systematic errors are well-known, for example, "unit errors" and "scanning errors" and are often identified and treated efficiently. Random errors have no systematic cause, for example, they may be due to source data suppliers making mistakes and require more sophisticated detection and treatment methods.

The methods used for detecting errors

checks for invalid and inconsistent data:

- requires identification of valid combinations

- inconsistencies within source and by comparison with others (for example, business register)

- duplicates and missing data

- negative values

- balancing errors

engage in plausibility checking of aggregate data:

- graphically

- through summary statistics

check for any suspicious patterns in data:

- identification of unusual values – "outliers"

- check with historic or time series data

- comparison with other sources

- the administrative source data holders may be aware of systematic problems with the dataset, which can help in identifying errors

- some standard editing techniques may be useful – testing is important

Office for National Statistics (ONS) has a choice in how to deal with suspicious values – the choice depends on time, resource, quality requirements and legal provision for querying data:

- ignore

- remove from dataset

- flag (potentially allow differential treatment for different uses)

- query with admin data holder or business

- manual imputation

- automatic imputation

5 . Administrative Data Quality Assurance Toolkit

The marks given in the following sections for risk, public interest, overall profile and assurance level have been determined using the UK Statistics Authority's [Administrative Data Quality Assurance Toolkit](). This toolkit is intended to help statistical assessors review the areas of practice for the quality assurance arrangements of administrative data used to produce official statistics. This guidance has been used to support critical judgements made by users of administrative data in making sure the source continues to be suitable for use in producing statistics.

6 . Quality assurance of administrative and business sources for the annual and quarterly national accounts

Office for National Statistics (ONS) has used the UK Statistics Authority's Quality Assurance of Administrative Data (QAAD) Toolkit to assess the risk of quality issues in its source administrative data. ONS has then looked at whether its methods for controlling those risks are appropriate to the risks posed. Using the notion of proportionality, ONS has looked at its areas of greatest risk and gauges the required detail with which it reports on those risks and how they are managed. On this basis ONS reports the following 12 risk areas with high levels of detail.

7 . Risk area number 1: VAT data from HMRC

Please note that the Value Added Tax (VAT) data source from HM Revenue and Customs (HMRC) is not currently used within the production of the output approach to gross domestic product (GDP). It is a source of data that is under development and the latest update on this project was published in February 2017. The assessment of VAT source data for the purposes of producing quarterly and annual national accounts is statistics of medium quality concern and higher public interest [A2/A3].

Quality (A2) – we regard the data as demonstrating a medium risk of data quality concerns when high risk factors have been moderated through using safeguards, for example, operational checks and effective communication arrangements. It is also appropriate to consider the extent of the contribution of the administrative data to the official statistics, for example, in cases where producers combine administrative data with other data types, such as survey or census data.

The public interest level of the statistics is high (A3). Quarterly and annual national accounts are economically important, reflected in market sensitivity; high political sensitivity, reflected by Select Committee hearings; substantial media coverage of policies and statistics.

Source and risk

Source: VAT data

Risk: Medium

Public interest: High

Overall profile: A2/A3

Assurance level: A2

Context and collection

HMRC makes this data available to Office for National Statistics (ONS) through monthly updates. There are two main variables: turnover and expenditure. These data supplement ONS's current Monthly Business Survey turnover data with HMRC turnover data using the strengths of both data sources. ONS runs a monthly process that allows the creation of microdata from the HMRC turnover dataset. The process allows matching of differing HMRC and ONS business reporting structures – the HMRC VAT reporting unit and the ONS reporting unit.

With access to VAT microdata, ONS has been able to construct "monthly" estimates from staggered quarterly VAT reports to overcome the issue of mainly quarterly data. This ensures that larger amounts of data could become available for estimates of gross domestic product (GDP).

ONS has provided an A2 enhanced level of assurance, which means that it has evaluated the administrative data quality assurance (QA) arrangements and published a fuller description of the assurance. ONS has provided users with a fuller description of the operational context and administrative data collection arrangements. It has:

- identified and summarised potential sources of bias and error in administrative system

- described safeguards taken to minimise risks to data quality

- described safeguards taken to minimise risks to data quality

Communication

ONS has provided an A2 enhanced level of assurance. ONS has agreed and documented:

- data requirements for statistical purposes

- legal basis for data supply

- data transfer process

- arrangements for data protection

- sign-off arrangements by data suppliers

ONS has established an effective mode of communication with contacts (for example, with HMRC, IT systems, operational or policy officials) to discuss the ongoing statistical needs in the data collection system and quality of supplied data, ONS has sought the views and experiences of statistics users and is resolving any quality issues.

Supplier quality assurance

In August 2013, the National Audit Office published an audit report on [The HMRC VAT service - the impact of legacy information and communication technology (ICT)](ADD LINK) {ADD LINK} in which they report the following.:

- The VAT ICT systems are robust and stable.

- HMRC has successfully extended the life of its legacy VAT ICT systems through enhancements.

- The VAT ICT systems are compliant with government security standards and have been independently assessed to accreditation requirements. Access to the systems is managed and controlled formally through the granting of appropriate user access rights and backup and recovery capability is regularly reviewed and tested.

- HMRC has a very experienced and knowledgeable internal team to support the VAT ICT systems. However, it will become increasingly difficult to source technical skills in the legacy technology (Virtual Machine Environment (VME) and COBOL) and to develop and retain expertise in the unique complexities and characteristics of the VAT ICT systems. HMRC recognises this risk and has embarked on an exercise to recruit a number of mainframe developers and implement succession plans and knowledge transfer activities.

- The hardware environment is provided through Aspire, HMRC's ICT supply contract with Capgemini and Fujitsu. Due to the scale, age and complexity of the HMRC systems only a small number of large ICT suppliers are able to support it and this will be a consideration when the current contract comes up for renewal in 2017.

- HMRC has a well-defined vision for ICT and for its business tax systems which incorporate VAT.

- HMRC completed a review of the end-to-end VAT process in 2012. One aim of the review was to reduce the degree of customisation and strive to standardise processes where possible. The core VAT processes have been mapped and a number of exception processes have been identified as potential targets for efficiency improvements.

- To reduce cost and complexity in line with strategic priorities, HMRC has implemented the Enterprise Tax Management Platform (ETMP), which will serve as a strategic ICT system for all business single taxes. While no firm plans to transfer VAT to ETMP are currently in place, we found that there is an aim across the ICT team to further explore the feasibility and potential benefits of a move.

- ETMP is based on modern technology and design principles and will represent a significant change from the existing VAT ICT systems. Therefore, to maximise the probability of achieving the full benefits of a potential move, the business process review team should be involved from the early stages to improve standardisation and reduce cost and complexity.

For performance information, we saw indications that HMRC has a good set of data that it uses in its day-to-day management.

HMRC Quality Report on its VAT statistics (2013) gives cursory details on the quality checks that HMRC conducts on the data. It does say that VAT cash receipts are the residual of HMRC's Indirect Taxes Consolidated Fund and the other Indirect Taxes. HMRC's Knowledge, Analysis and Intelligence Directorate (KAI) receives data on HMRC's individual bank accounts for indirect taxes on a daily basis. From other sources, KAI compiles and matches receipts from all other indirect taxes. They are then subtracted from the total amount in the Indirect Taxes Consolidated Fund at the end of each month, the residual being categorised as VAT. KAI also uses the Customs and Excise Core Accounting System (CECAS) to obtain data on VAT paid on imports, which is administered separately.

ONS has provided an A1 basic level of assurance as it has reviewed and published a summary of the administrative data QA arrangements. ONS knows that HMRC conducts QA checks but has not published a description. ONS knows that audits are conducted on the admin data but has not described the implications for the statistics.

Investigations and documents

An advantage of the combined ONS survey and HMRC turnover approach is the ability to validate and quality assure data between the two data sources.

Unit errors are common in survey data but they are also present in VAT data. In the UK, businesses should report VAT in pounds, some businesses report in thousands of pounds, by habit. These errors are easy to correct, if the national accountants can identify them. The most effective way of identifying them is by comparison with previous data. ONS adopts a thousand pound rule, which works by calculating a ratio between current VAT return in comparison with a company's previous VAT return. If the ratio falls between the values 0.00065 and 0.00135 it is multiplied by 1,000.

Scanning errors may occur during data entry. Often statisticians can identify them as highly implausible values. The system may use special characters to denote scanning error. For UK VAT data, they are set to 99999999999 but this does not necessarily capture every scanning error. It is simple to identify a pre-set code but it may need additional edits for other scanning errors.

ONS looks for suspicious quarterly data patterns and applies a quarterly patterns rule, but only applies this rule to businesses who report their VAT on a quarterly basis. There are three variations for the quarterly patterns rule. The aim is to try to understand whether these businesses are reporting true quarterly HMRC turnover, by identifying suspicious quarterly patterns. Therefore, the cleaning under this rule is based on:

- reporting units having exactly the same positive values for any 4 consecutive quarters; this implies that the business is actually reporting annual values allocated equally between the 4 quarters

- reporting units having exactly the same positive values in any 3 consecutive quarters and then a different value for the fourth quarter; this implies that the business is assessing its annual value and allocating it between the 4 quarters; the fourth quarter therefore is allocated the residual value to sum to the annual value

- reporting units having zero values in any 3 quarters and then a positive value in the fourth quarter; this implies the business is returning an annual value

It is fairly simple to identify patterns, where one period looks incorrect, ONS imputes using its preferred method for a single erroneous or missing value. ONS does not treat negative values by simply reversing the sign, instead they impute a new value. Where ONS suspects an annual figure, they re-apportion the annual figure amongst the quarters. ONS tested the methods by randomly creating suspicious patterns in clean data and comparing their imputations with true value. This led to the method of apportionment that proved to be the most accurate.

The suspicious turnover rule identifies reporting units that have suspicious turnover for a VAT return. ONS deems a return as suspicious by firstly matching a company's current VAT return to its previous VAT return. (This applies for all the reporting schedules.) When ONS deems a value as suspicious as a reporting unit it is compared with all reporting units within that employment size-band at the UK Standard Industrial Classification: SIC 2007 class level. (The employment size-bands have been set at employment groupings of 0 to 9, 10 to 49, 50 to 99 and 100 and over.)

Once the data has been stratified by class and employment size-band the reporting unit's current and previous VAT returns will be tested in comparison with the median value of the class and reporting stagger. A set of criteria in terms of scores is then produced and if it falls outside these scores it will be deemed as a suspicious turnover value. This is then replaced by a value which is the ratio of the current period sum of VAT turnover divided by previous period sum of VAT turnover for the total UK SIC 2007 class and employment size-band multiplied by the reporting unit previous period VAT turnover figure. A clean marker is then applied to the reporting unit to ascertain whether the data has been cleaned or not ("0"indicates no cleaning applied "1 to 5" indicates cleaning has taken place on the reporting unit).

The amendment of values due to the suspicious turnover rule is an area where ONS intends to do further work as at present some of the values are not cleaned to a credible path. So at present ONS has decided to be consistent in only using and sharing data that have NOT been cleaned using these three rules rather than adopting two cleaning practices while omitting the third.

A degree of manual editing is required after the impact of automatic cleaning to ensure that the final results are of an appropriate quality. An initial view from manually cleaning five of the initial candidate industries being used to pilot the approach suggests that few businesses require manual editing but where these occur the impact can be significant. As of April 2016, ONS is confident in identifying businesses that require intervention in this way and in its ability to improve the automated cleaning rules and seek to limit the ongoing need for manual editing.

As businesses cannot be contacted by ONS to query unusual data points it's not possible to find the truth – this makes evaluation of methods difficult. Diagnostics have been proved helpful, identifying estimated proportion of suspicious businesses, the average size of businesses identified as suspicious and the average VAT values of suspicious businesses

To correct random errors in VAT data imputation has been found to be the only option for "correcting" random errors. A range of imputation methods were tested and the best method selected.

ONS has provided an A3 comprehensive level of assurance as it investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit. ONS has:

- provided a detailed description of its own QA checks on the admin data (including validation, sense and consistency checks) given quantitative (and where appropriate qualitative) metrics for specific quality indicators

- undertaken comparisons with other relevant data sources (such as survey or other admin data)

- identified the strengths and limitations of the admin data and any constraints on use for producing statistics explained the likely degree of risk to the quality of the admin data

8 . Risk area number 2: Data used to calculate FISIM

The assessment of monetary financial institutions (MFI) source data from the Bank of England (BoE) for the purposes of calculating financial intermediation services indirectly measured (FISIM) as a component in producing quarterly and annual national accounts is statistics of medium quality concern and higher public interest [A2/A3].

Quality (A2) – the data may be regarded as having a medium risk of data quality concerns when high risk factors have been moderated through using safeguards, for example, operational checks, and effective communication arrangements. It is also appropriate to consider the extent of the contribution of the administrative data to the official statistics, for example, in cases where the statistics are produced in combination with other data types, such as survey or census data. There is no other data that can be used to triangulate calculations of FISIM so the Bank data is single source heightening the risk.

The public interest level of the statistics is high (A3). Quarterly and annual national accounts are economically important, reflected in market sensitivity; high political sensitivity, reflected by Select Committee hearings; substantial media coverage of policies and statistics.

Source and risk

Context and collection

Banks and other financial intermediaries earn income not only through explicit fees and charges but also on the margin between the interest they pay to depositors and the interest they charge borrowers.

Because of the margin, the financial services consumed do not necessarily have an explicit price, unlike most goods and services in an economy.

For the purposes of the national accounts the services provided, through the use of the margin, are known as Financial Intermediation Services Indirectly Measured or FISIM. The fees earned by financial institutions are quite easy to measure. Banks make explicit charges for some services, such as commission on foreign exchange, account charges and flat rate fees for overdrafts. However, the amount of these charges is significantly below the costs paid by the banking industry on wages, bonuses and other intermediate costs (for example, office space, computer equipment, travel).

Intermediation services on loans and deposits are difficult to measure but forms approximately 66% of total implicit and explicit financial services charges (or 66% of output) for the banking sector in 2005.

Calculating the value of a service without a price is complex. However, the international System of National Accounts (SNA) provides guidelines on using a relevant reference rate to achieve this. In addition to how the value of FISIM is calculated, what is important for gross domestic product (GDP) estimates is how statisticians allocate those financial services to their end-users.

Within the SNA framework, goods and services consumed by businesses in order to produce other goods and services, are treated as intermediate consumption (or an expense) and subtracted from GDP. However, the use of goods and services by household, government and the foreign sector are treated as final demand and these add to GDP.

This approach means that FISIM is having two separate effects within GDP:

- a net effect on GDP: it increases GDP by 1 to 2.3% (it added £16.6 billion or positive 1.1% to GDP in 2011)

- a redistribution effect: it increases the contribution of financial services to GDP (positive £70 billion in 2011) and reduces the contribution of all other sectors to total GDP (negative £54 billion in 2011)

Office for National Statistics (ONS) has provided an A1 basic level of assurance, which means that it has reviewed and published a summary of the administrative data quality assurance (QA) arrangements. ONS has:

- provided users with an outline of the administrative data collection process

- outlined the operational context

ONS has not identified actions taken to minimise risks to quality nor identified and summarised the implications for accuracy and quality of data, including the impact of any changes in the context or collection arrangements.

Communication

BoE uses the latest version of the published ONS Sector Classification Guide for counterparties to identify their source public sector banks.

There is a broad coverage of sectors included on the statistical returns. Some members of the public banking groups may report on a quarterly basis but these make up only 5% of the monetary financial institutions (MFI) population. Monthly data is imputed for quarterly reporters.

Revisions are routinely taken on from reporters and published monthly. Other methodological revisions will be delivered in agreement with ONS.

Form PL collects the profit and loss of UK MFIs. Data are disaggregated into main components, for example, interest, fees, dividends, trading income, other operating income and expenditure.

Data are taken from a sample of those institutions licensed by the Prudential Regulation Authority (PRA) to accept deposits. The sample is 95% quarterly and 98% annually. The remaining 2% of business is modelled.

For the public sector finance dataset we use reported information by those banks that come within the boundary of the public sector.

A firm agreement between the Bank of England and ONS sets out that if the necessary resources can be made available, the Bank agrees to carry out consultancy work in areas where it has responsibility for data collection. The specification and targets for this consultancy work, which will be focused on developing sources and methods for the coverage of innovation and on improving coverage of existing financial products and institutions, will form the basis of the Bank's joint work programme with ONS. The work programme will be agreed by both organisations and both organisations will seek to complete it as quickly as resources permit.

The firm agreement sets out that any changes to the Bank's procedures in compiling the figures, to the extent that they affect the quality or relevance of the data, shall wherever practicable be discussed with ONS before they are made. In addition, the Bank and ONS will liaise closely on their response to evolving needs for changes to the collection systems through which the data are obtained.

Arrangements for monitoring and review

ONS and the Bank monitor the operation of their firm agreement and have undertaken to inform each other at an early stage of any emerging problems.

There are quarterly reports on the delivery of data and adherence to timetables will also be monitored by both sides and perceived issues raised in the first instance at quarterly liaison meetings. A formal annual assessment of the Bank's success in meeting its obligations under the Agreement, covering periods to end-March, will be completed by ONS within 3 months of that date and a meeting of signatories and chief contact officers will be convened shortly thereafter to take forward issues arising from the assessment.

Note that we are unable to assess this currently since ONS has not published a summary of the administrative data QA arrangement. There is no evidence currently that ONS has agreed and documented:

- data requirements for statistical purposes

- legal basis for data supply

- data transfer process

- arrangements for data protection

- sign-off arrangements by data suppliers

It is anticipated that ONS has established an effective mode of communication with contacts (for example, with the Bank of England, IT systems, operational or policy officials) to discuss the ongoing statistical needs in the data collection system and quality of supplied data.

We have yet to confirm whether ONS has sought the views or experiences of statistics users and resolved any quality issues related to FISIM.

Supplier quality assurance

Banks provide data on their total loans and deposits to the Bank of England on a quarterly basis, which are then aggregated and passed to ONS for inclusion in output statistics. The Bank collects a range of administrative data (including regulatory data) as well as statistical data.

The International Monetary Fund (IMF) has developed a Data Quality Assessment Framework (DQAF) arising out of its work in promoting the General Data Dissemination System and Special Data Dissemination Standard (SDDS). The current version of the DQAF dates from 2012, and covers both generic and domain-specific statistical standards. It addresses "quality" in a very broad sense, including topics such as relevance, integrity and privileged access, which are covered in other sections of the Bank's Code beside the specific definition given previously. The UK has subscribed to the SDDS since its inception in 1996 and this carries a commitment to meet the standards set out in the DQAF.

The statistical domains covered include series published both by ONS and by the Bank.

An SDDS subscriber has to submit information about its data, its dissemination practices and its metadata to the IMF, for review of comprehensiveness and international comparability.

National metadata are published by the IMF, using a standard template on its website as part of the Dissemination Standards Bulletin Board.

Similarly, the European Central Bank (ECB) has promulgated a public commitment on European Statistics by the European System of Central Banks (ESCB) drawing on the same underlying sources and setting equivalent standards. The UK is not a member of the euro area and the Bank is not subject to the statistical reporting requirements of the ECB, but the frameworks for data quality follow similar approaches.

The present Bank Data Quality Framework adopts the ESS dimensions of statistical data quality, in order to address the requirement in the Bank's Statistical Code of Practice that the Statistics Division should progressively develop and publish quality standards for its statistics.

The ESS standards have been adopted in the UK by ONS as the basis for data quality measurement and reporting set out in their "Guidelines" framework. That framework comprises a core set of 11 main quality measures and a larger number of other quality measures, quality indicators and guidelines. The main quality measures are prioritised as the minimal quality data set for all ONS statistical outputs where they are relevant.

Central banks often operate under distinct legal powers for data gathering, applicable to all entities in specified banking and other financial subsectors of the economy. In the UK, Section 17 of the Bank of England Act 1998 grants power to require provision of information about relevant financial affairs that the Bank of England considers it necessary or expedient to have for the purpose of its monetary policy functions from all banks and building societies, and certain other categories of institutions (including financial holding companies, issuers of debt securities and arrangers or managers of such issues). The use of this power enables the Bank to publish monetary and financial statistics in the public interest.

Because of the complexity of the full data set of required monetary and financial statistics items, and the relatively small population, the accepted view has been that to use random or stratified random sampling methods (under which reporters would rotate into and out of samples over time) would be an onerous and impractical proposition. Such methods have therefore not been adopted.

Statistical information to be reported to the Bank is set out in a number of standardised returns or reporting forms. All such forms, their guidance notes and validation rules, are published on the Bank's website. Statistical returns are organised in a systematic format appropriate to the information focus of the particular return. Individual data items including any currency, sectoral or other classification splits and relevant aggregation subtotals, are mapped to coded boxes on the return. Statistical returns have specified reporting panels and submission deadlines. Not all reporting institutions need to report all statistical returns; at a minimum, however, all monetary financial institutions report the core balance sheet return, Form BT, at a quarterly frequency. Electronic submission of reporting forms is encouraged.

The consequence of this system is that in practice most Bank statistical outputs are generated from a near-census coverage of the bank and building society sectors. For UK monetary aggregates data, coverage was 98% of monetary financial institutions by balance sheet size (size of total assets), at a monthly frequency of reporting at the end of 2013, and there was full coverage at quarterly frequency.

There is a similar situation in the euro area, where under current rules the European Central Bank requires national central banks to maintain a minimum 95% coverage of monetary financial institutions by balance sheet size.

Where less than a full census, coverage is usually determined by the application of minimum reporting thresholds: all institutions above specified thresholds are required to report statistical returns. Thresholds are usually based on balance sheet size, or on criteria specific to the information collected on the return, and are periodically reassessed and announced. The objective in general terms is to minimise reporting burdens, subject to maintaining sufficient coverage of the reporting population so as to yield accurate estimates of the data. Determining a reporting population on a threshold basis is termed "cut off the tail" sampling, distinguished from random or stratified random sampling.

Statistical reporting by banks and building societies can sustain a much greater complexity of data reporting than it would be reasonable for ONS to impose in its household and business surveys. In major banks, responsibility for statistical reporting is typically located in a financial and regulatory reporting unit. Such arrangements might be presumed to deliver a degree of quality assurance, as well as process economies for the reporters, which would be less likely to arise for typical entities of comparable size in other economic sectors.

Currently this is an A0 level – ONS does not do any further checks on the lower level data due to the quality and detailed checks carried out by the Bank. Although ONS applies an A0 level to this data source, it did find an A2 enhanced level of assurance from details provided publicly by the Bank itself.

Also, the firm agreement between the Bank of England and ONS sets out that the Bank and ONS recognise the importance of producing accurate data, within an appropriate conceptual framework. To ensure that data are fit for purpose, the quality of the data is assessed regularly, and improvements are made where necessary. The Bank will take all reasonable steps to achieve satisfactory and representative coverage of institutions measured by its surveys. Where figures are supplied to ONS based on survey data and all inquiry forms have not been received (that is, where imputation is materially greater than usual) this will be made known to ONS at the time the figures are provided.

ONS and the Bank undertake before the end of each calendar year to review the basis of the timetables for the supply of data between them and, in respect of data exported via the SRDD Outputs System, to produce an agreed detailed timetable covering the succeeding calendar year. Both parties undertake to meet the timetables and to give reasons if on any occasion this undertaking cannot be met. Changes to any timetable may be agreed at working level between the two organisations at any time, and the chief contact officers informed.

The Bank will supply data amendments for previous periods with data for the current period and will provide ONS at that time with a copy of its internal notes on revisions. Where these revisions have a potential material effect on published national accounts and balance of payments aggregates, and to the extent that the reasons for such revisions are not clearly apparent from the Bank's internal notes, ONS may request explanations.

Investigations and documents

There are certain differences in approach in the treatment of identified or suspected data reporting errors revealed by data cleansing.

ONS's approach is a mixture of follow-up with respondents for major discrepancies, acceptance of unusual values (which cannot reliably be distinguished from reporting errors) if there is some supporting evidence and editing (that is, replacing with an imputed value) the data stored in their internal database system. ONS also keeps both the edited and unedited microdata values within these systems, to facilitate evaluation of the process and tuning of the editing parameters.

The Bank's approach is to require reporters to resubmit corrected data, but not otherwise to edit individual bank inputs at the raw data level. Where there are known or suspected reporting errors but corrected data inputs have not been received from reporters, there exists an "adjustments" facility by which estimated corrections can be applied to any consequential statistical outputs.

It should be noted, however, that the adjustments facility is used more routinely to account for revaluation effects and for remaining "other changes in the volume of assets" (OCVA) such as write-offs or population changes, in the estimation of financial flow data series. It follows that, unlike the definition of the editing rate as one of ONS's main quality measures, application of the adjustments facility in the Bank's system cannot be interpreted as a pure indicator of data errors.

Data quality differences also arise in the final stages of production of statistical outputs. The majority of monetary and financial statistical data series are expressed at current prices and in non-indexed units; for these data, issues of index construction do not arise. National accounts data, on the other hand, are normally expressed in both current price and constant price terms, and under annual chain-linking, constant price indices are re-based each year. National accounts estimates may be subject to balancing adjustments in order to ensure coherence of estimates produced from independent survey sources. In these respects, monetary and financial statistics are naturally subject to fewer causes of revision than is the case for national accounts data.

As well as for monetary policy purposes, the Bank of England has a parallel statutory power to collect data from banks and building societies in order to compute statutory cash ratio deposits (CRDs), being part of the mechanism by which the Bank is funded. CRDs are compulsory non-interest bearing deposits held at the Bank of England, set for each liable institution as a specified fraction of their eligible liabilities above a certain threshold. A high coverage of institutions is required to ensure that all banks above the CRD threshold are included.

Another distinction relates to the reliability of data inputs reported to the central bank. The Bank reports that there is frequent two-way communication between reporters (individually and collectively) and its statistical staff, as well as a culture of high standards of statistical reporting by monetary financial institutions. More recently, reporting has been on a statutory basis since 1998 so that, formally, failure to supply requested information and knowingly or recklessly to provide information "which is false or misleading in a material particular" are defined as offences subject to penalties, under Sections 38 and 39 of the Bank of England Act. In practice, the Bank expects that written warnings would normally be sufficient for problems of poor reporting to be addressed.

Non-response rates are typically zero for Bank reporters, so this source of data error is largely absent. By contrast, in the case of household and business sector surveys conducted by ONS and other national statistical institutes, these imperatives may be less strong and less easily enforceable, so that instances of sizable non-response rates may arise.

Measurement of FISIM in aggregate is by observing margins. The appropriateness of including the cost of wholesale funds in margins and of debtor or creditor approaches for the measurement of bank interest , as well as the impact the bad debt write off, write down and write back on FISIM calculations are all material to the quality of the statistics.

Processes and quality assurance checks are reviewed continually each month and data are consistent with European System of Accounts: ESA10 guidelines.

They continually check for unusual movements in the data compared with the past history of that data point, large movements contributing to aggregate measures; asking reporters for counterparties to check their classification and business reasons to understand the context of movements. Reporting institutions can access sample forms, definitions and validations for statistical returns, which are available on the Bank of England's external website. Whenever changes are made to such documentation, a Statistical Notice is published to inform reporters of how this impacts their reporting requirements. These documents are available on the Bank of England's website.

The specification for the series was drawn up by ONS using BoE data sources and this provides the framework for the public sector data provided to ONS.

Checks are made to ensure consistency with known events e.g. the sale of 4G licences. Interbank differences are followed up with the MFIs concerned.

Processes and quality assurances are reviewed on an ongoing basis and data are consistent with ESA10 guidelines.

All reported data are initially checked for plausibility at an individual institution level. Any unusual movements in the data are queried with the reporting institution that then provides a business explanation for the movement. If the explanation is felt insufficient, further queries will be raised with the institution until they are satisfied they have an explanation and the reporting is correct. If they identify a case of misreporting, the institution will be asked to resubmit a corrected form.

Aggregate data are also reviewed and any large unchecked movements are followed up in the same vein as previously.

Reporting institutions submit revisions as and when they are identified. Any potentially large revisions from one institution are managed in line with national accounts' revisions policy.

Given the level and detail of BoE checks, ONS does not make any further Q&A checks on the lower level data. Aggregate level data are checked as part of the regular compilation process.

9 . Risk area number 3: Central government expenditure data

Source and risk

Source: HM Treasury's (HMT's) public spending database, OSCAR (Online System for Central Accounting and Reporting)

Quality risk: Medium

Public interest: High

Overall profile: A2

Assurance level: A2

Context and collection

Expenditure data are entered as accrued values in accordance with the UK Fiscal Framework. Precise accounting rules that government departments must follow are set out in the Financial Reporting Manual (FReM).

Spending data from OSCAR are modified with a number of adjustments to meet national accounts requirements.

Communication

HMT publishes raw data from OSCAR quarterly.

While Office for National Statistics (ONS) is ultimately responsible for the quality of the public finance data it uses, the same data source is used by a number of government departments for their own planning purposes, who therefore have an interest in ensuring the quality of the data.

OSCAR data is used by HMT for its own fiscal planning purposes and by the Office for Budget Responsibility (OBR) for forecasting the government finances.

Copies of summary checks completed by ONS are sent to HMT officials who compare these with ones produced by their own systems. Any differences are highlighted and the reasons identified. If necessary the data are corrected and the systems re-run.

Supplier quality assurance

Statisticians in HMT ensure that OSCAR maintains its integrity as a data source.

Departmental performance in supplying monthly data is assessed according to a range of qualitative and quantitative indicators. Departments are given feedback on their performance. Summary data from OSCAR are provided monthly by HMT to ONS, and a detailed extract provided every quarter. In addition OSCAR data are made publicly available every quarter.

Investigations and documents

Govt expenditure data goes through various stages of refinement during the financial year and beyond.

Detailed central government expenditure data from OSCAR are received by an ONS team who aggregate the data into low--level quarterly time series, which can be aggregated together later to calculate the higher-level central government fiscal statistics required by national accounts. The calculated data series typically identify: the European System of Accounts (ESA) transactional areas; the Classification of the Functions of Government (COFOG) category, the counterpart sector as well as more specific information about the nature of the transaction.

Data are processed by ESA transactional area using SAS and Excel. The data are loaded to the ONS relational database system and further tasks and consistency checks run within this system. Summary reports are produced from an ONS system and checks for further quality assurance are made, based mainly upon revision and growth rate analysis.

Adjustments are made for a number of reasons:

- conceptual framework differences

- error corrections – sometimes there are errors in the data that departments load onto OSCAR; although every effort is made to correct the data on the database itself, in some cases this is not possible within the tight publication timetable and adjustments are made on top of the data instead

- use of sources, other than OSCAR – adjustments are necessary where ONS uses data sources other than OSCAR,, for example, depreciation where ONS uses the perpetual inventory model (PIM) rather than a department's own estimates

10 . Risk area number 4: Government gilt and Treasury Bill data by auction and redemption

Source and risk

Source: Debt Management Office Outputs: Blue Book, Quarterly national accounts and UK Economic Accounts

ONS compiler area: Public sector finance - Fraser Munro

Quality risk: Medium

Public interest: High

Overall profile: A3

Assurance level: A3

Context and collection

The primary purpose for which Debt Management Office (DMO) collects these data is so that they can carry out the government's debt management policy, which includes minimising financing costs over the long-term, taking account of risk and minimising the cost of offsetting the government's net cash flows over time.

DMO holds weekly tenders as part of its debt and cash management activities offering 1-month, 3-month and 6-month maturities. The tender is run electronically using the Bloomberg Auction System.

Primary participants (banks eligible to bid at the tenders) place their bids through the Bloomberg Auction System. When the tender closes at 11am, these bid data are then transferred electronically from the Bloomberg Auction System to the System for Electronic Auction and Tender (SEAT), the UK's DMO auction platform. The SEAT system records all auction and tender transactions that take place, therefore there is no loss in coverage and no source of bias in the data collected.

Treasury Bill auction results

A summary report of Treasury Bill Tender (TAS007) results for each of the maturities offered at tender, that is, 1-month, 3-month and 6-month, is created from the SEAT system and sent to ONS on a monthly basis. This report presents the following data for each maturity:

- lowest accepted yield
- highest accepted yield
- average rate of discount
- average price per £100 nominal
- tail (in yield terms)
- tap rate
- amount tendered for (£)
- amount on offer (£)
- cover
- residual amount (£)
- amount allocated (£)
- cost (£)
- discount (£)

This summary report covers each of the tenders held in the relevant month (4 or 5 tenders).

Gilt auction results

Gilt auction data is provided in three deliveries:

- the announcement; primarily indicating the gilt to be auctioned and the date of the auction

- the result; primarily indicating the auction price

- the post-auction offering; primarily indicating any additional quantity of the gilt offered (up to 10% of the auction) offered at the final auction price

Treasury Bill portfolio composition

This Excel release provides a point of quality assurance for the quantity of Treasury Bills auctioned by type in any month, along with supplementary data not available via auction results (such as additional Treasury Bill operations carried out by DMO) and a check of the total Treasury Bill stock outstanding at the end of any month.

DMO quarterly review

This publication lists the total Treasury Bill and gilt stock outstanding each quarter. It serves as an additional quality assurance check for the Public Sector Finances (PSF) Branch by allowing us to compare the numbers we produce with the numbers published in this report.

We believe DMO's data collection processes and the sharing of information between ONS and DMO meet the A3: Comprehensive assurance level required.

11 . Risk area number 5: Brokerage for chartering, sales and purchases of ships and aircraft

Source and risk

Source: Baltic Exchange

Outputs: The data are used in calculating total trade in financial services of the quarterly national accounts and the second estimate of gross domestic product (GDP)

ONS compiler area: UK trade - Katherine Kent

Quality risk: Medium

Public interest: Medium

Overall profile: A2

Assurance level: A2

Context and collection

This data makes up a small component of total trade in financial services (less than 2%, the majority coming from surveys).

Data is provided voluntarily by members of the Baltic Exchange organisation. Around 170 members are sent a broker survey questionnaire. Normally around 40% to 45% respond and it is ensured the top 10 biggest companies always respond.

Imputation is used to fill in for non-responders, the size of company is taken into account.

Note: It is commercial market data used by ONS under licence.

Communication

Single point of contact identified, further contacts provided. Future teleconferences arranged.

Supplier quality assurance

Good knowledge of processing in Excel and keen to start survey with support from ONS.

Investigations and documents

Quarter-on-quarter, year-on-year and quarterly-to-annual consistency checking.

12 . Risk area number 6: Air and surface mail expenditure and receipt

Source and risk

Source: Royal Mail Group

ONS compiler area: UK trade – Hannah Finselbach

Risk: Medium

Public interest: Low

Overall profile: A2

Assurance level: A2

Context and collection

Air transport services – Royal Mail Group feeds into services and transport air imports freight – other imports comprise payments to non-resident airlines for carrying UK airmails.

Contributes to other air transport revenue, which is a historically low figure. Also contributes to postal and courier services under communication services.

One individual supplier contributes a very small proportion to the overall figures that are published in the annual Pink Book and the quarterly national accounts.

Monthly, quarterly and annual data supplies, the majority of which are voluntary. The data covers sea, air, road and other transport services.

Although there is a mixture of medium and high concerns over data quality, it is important to note that this is offset by the small contribution that each of the outputs carry towards the overall output of transport services.

Communication

Initial meetings held and single point of contact identified with further contacts acquired.

Further teleconferences arranged on an annual basis

Supplier quality assurance

Strong quality assurance and regular meetings with their suppliers and benchmarking is carried out. Commentary is provided on movements. Good response rates.

Investigations and documents

Quarter-on-quarter, yearon-year and quarterly-to-annual consistency checking. Background quality information provided in most cases:

- data are usually received on time on a quarterly basis; however, a recent freeze on using all Royal Mail data has caused ONS to revert to estimates.

- data filters through a variety of spreadsheets, increasing chances of keying errors; internal project being carried out to look at reducing the number of spreadsheets

- responses to our queries are generally well received from supplier

- voluntary agreement to provide data.

- no knowledge on how the data is collated or expertise of the compiler

- trying to ascertain new sources of data, discussions on whether this could be sourced from another ONS survey

13 . Risk area number 7: Benefits abroad

Source and risk

Source: HM Treasury

Outputs: Quarterly national accounts and second estimate of GDP

ONS compiler area: UK trade, Hannah Finselbach

Risk: Medium

Public interest: Medium

Overall profile: A2

Assurance level: A2

Context and collection

Government services have a medium risk level of quality concern, there are three data sources and all have been assessed at medium risk, this is based on the following initial observations:

- source data contributes a very small weighting to the overall government services figures

- limited knowledge on how the data is collated or the expertise of compiler

- data are supplied on a voluntary basis

Government services includes: all transactions by embassies, consulates, military units and defence agencies with residents of staff or military personnel in the economies in which they are located.

Communication

Overall assurance level A2. To mitigate this level of risk we have carried out the following:

- regular contributions received from supplier

- consistency checks are run in ONS systems and quarter-on-quarter and year-on-year movements are carried out internally

- face-to-face meeting arranged to ascertain the strengths and weaknesses of the data and to establish quality assurance measures

- quarterly meetings to be carried out to maintain supplier relationship

Supplier quality assurance

Possible issue over timing but considered negligible.

Strong quality assurance – auditing carried out weekly if not daily.

Investigations and documents

Regular and timely contributions received from supplier; limited knowledge on how the data is collated or the expertise of the compiler, discussions are ongoing to ascertain this.

Consistency checks are run in our systems and quarter-on-quarter and year-on-year movements are carried out internally.

Data are supplied on a voluntary basis.

14 . Risk area number 8: Accident and health income and outgoing: property claims data

Source and risk

Source: Association of British Insurers (ABI)

Outputs: Blue Book, Input-output analytical tables and Quarterly national accounts

ONS compiler area: Household expenditure - Gareth Powell

Risk: Low

Public interest: Low

Overall profile: A1

Assurance level: A2

Context and collection

Source data makes up a small percentage of overall household final consumption expenditure.

Some knowledge of the steps taken to compile data and expertise of the organisation compiling it.

Data available through a membership.

Communication

M2 – enhanced level of assurance. Quarterly and annual deliveries received from supplier.

Queries are raised with supplier in the event of issues regarding timeliness and quality of deliveries.

Data accessed through password protected membership. Consistency checks run in CORD.

Supplier quality assurance

Timeliness is sometimes an issue, which occasionally requires the supplier to be contacted.

No knowledge of supplier's own quality assurance processes.

Investigations and documents

Quality assurance checks carried out and contact made to the supplier when necessary.

Consistency checks run in CORD.

Data accessed through password protected online membership..

15 . Risk area number 9: Construction, design and management regulations (CDM) information

Source and risk

Source: Scottish Water

Outputs: Blue Book, Input-output analytical tables, Quarterly national accounts and Second estimate of GDP

ONS compiler area: Household expenditure - Gareth Powell

Risk: Low

Public interest: Low

Overall profile: A1

Assurance Level: A1

Context and collection

Scottish Water have a low risk level of quality concern. Source data contributes a very small percentage of overall household final consumption.

Some knowledge of how the data is compiled and the supplier's expertise.

Data supplied voluntarily upon request.

Scottish Water covers every household and building making use of their services.

Communication

Annual delivery emailed to team upon request.

Liaison contact established with supplier to raise any queries.

Supplier quality assurance

No knowledge of supplier's own quality assurance processes.

Rarely issues with timing.

Investigations and documents

Quality assurance checks carried out and contact made to the supplier when necessary.

Consistency checks run in CORD.

Data supplied voluntarily upon request.

16 . Risk area number 10: Premium Bonds and net financing of National Savings and Investments

Source and risk

Source: National Savings and Investments

Outputs: Blue Book, Quarterly national accounts and UK Economic Accounts

ONS compiler area: Public sector finances - Fraser Munro

Risk: Low

Public interest: Low

Overall profile: A1

Assurance level: A1

Context and collection

There are a number of small amounts of interest that the government pays out and are scored to public sector net borrowing as they accrue. These include interest on National Savings certificates.

A small number of expenditure items come from sources other than OSCAR.

The largest of these, expenditure on debt interest, is computed from a variety of sources – including National Savings and Investments.

Communication

Unknown

Supplier quality assurance

Unknown

Investigations and documents

Unknown

17 . Risk area number 11: Residential telecommunication activity

Source and risk Source: Ofcom

Outputs: Blue Book, Input-output analytical tables, Quarterly national accounts

ONS compiler area: Household expenditure - Gareth Powell

Risk: Low

Public interest: Low

Overall profile: A1

Assurance level: A1

Context and collection

Source data makes up a small percentage of overall household final consumption expenditure.

No knowledge of the steps taken to compile source data.

Some knowledge of the expertise of compiler organisation.

Data publically available online.

Communication

M1 – Basic level of assurance. Data obtained from website each quarter.

Queries raised with supplier in the event of issues with data quality.

Data publically available.

Consistency checks run in CORD.

Supplier quality assurance

No knowledge of supplier's own quality assurance processes. Data is quality assured by branch upon receipt.

Investigations and documents

Quality assurance checks carried out and contact made to the supplier when necessary.

This has a minimal impact upon output.

18 . Risk area number 12: National Lottery data

Source and risk

Source: Ofcom

Outputs: Blue Book, Input-output analytical tables, Quarterly national accounts

ONS compiler area: Household expenditure - Gareth Powell

Risk: Low

Public interest: Medium

Overall profile: A2

Assurance level: A1

Context and collection

Source data makes up a small percentage of overall household final consumption expenditure.

Some knowledge of the steps taken to compile source data.

Some knowledge of the expertise of compiler organisation.

Data publically available online.

Communication

M1 – Basic level of assurance.

Data obtained from website each quarter for sales, twice a year for data on prizes.

Queries raised with supplier in the event of issues with data quality.

Data publically available.

Consistency checks run in CORD.

Supplier quality assurance

No knowledge of supplier's own quality assurance processes. Data is regulated by National Lottery Commission, which is part of the Gambling Commission. Data is quality assured by branch upon receipt.

Investigations and documents

Quality assurance checks carried out and contact made to the supplier when necessary.

This has a minimal impact upon output.