Statistics and Regulatory Data Division

Data Quality Framework

This Data Quality Framework is designed to enable users of the Bank’s\(^1\) published statistical data to be better informed about aspects of the quality of those data. Information on data quality can have a wide scope: it consists of explanatory material describing the relevance of data, how statistics are compiled, the use of imputation methods, and any other information (including quantitative measures) useful to users in their understanding of what the data represent and how they are constructed. The Framework includes a number of references to explanatory articles relevant to data quality previously published by the Bank’s Statistics Division\(^2\) and other bodies. It describes how the Bank’s approaches to statistical data quality conform to international standards on quality assessment. Quantitative indicators relating to the coverage and revisions aspects of data quality are defined.

In April 2013 the Statistics Division took over responsibility for the production of a wide range of regulatory data on banks, building societies, insurance companies, investment firms, credit unions and friendly societies previously undertaken by the Financial Services Authority. At present, these data collections result in published statistical outputs only to a very limited extent. This Framework applies to those outputs; but the Statistics Division is in the course of developing a similar quality assurance Framework for the regulatory data it processes for the Bank’s own and other official uses.

We intend to review this Framework at least every five years, issuing a revised version whenever material changes are required.

In June 2013 the Bank published a revised version of its Statistical Code of Practice,\(^3\) following the regular triennial review.\(^4\) The Code sets out standards for the collection, compilation and dissemination of data by the Bank for statistical and regulatory data purposes. It is aimed at ensuring quality, objectivity, openness and freedom from political interference. It rests upon seven key principles: relevance, integrity, quality, accessibility, confidentiality, respondent burden and cost efficiency.

---

\(^{1}\) Throughout this Framework, the convention has been adopted that reference to ‘the Bank’ applies jointly and equally to the Governor and Company of the Bank of England and the Prudential Regulation Authority as its subsidiary. Where the context specifically requires separate consideration, ‘the Bank of England’ or ‘the PRA’ is used.

\(^{2}\) The Statistics and Regulatory Data Division of the Bank was formerly known as the Monetary and Financial Statistics Division; its name was changed in 2013 to reflect its expanded scope of responsibilities. The convenient abbreviation ‘the Statistics Division’ is used throughout this Framework.

\(^{3}\) The revised version of the Code and an accompanying article explaining why revisions were needed and summarising the main changes, particularly those arising from the Financial Services Act 2012, can be viewed here: www.bankofengland.co.uk/statistics/pages/about/code.aspx. The Financial Services Act 2012 can be viewed here: www.legislation.gov.uk/ukpga/2012/21/contents/enacted.

\(^{4}\) In the Code we commit to undertake regular reviews of its content. Currently these take place triennially.
The Bank welcomes questions and requests for further information on this Data Quality Framework, via email at SRDD_Editor@bankofengland.co.uk or via post addressed to:

Business and Communications Group
Statistics and Regulatory Data Division
Bank of England
Threadneedle Street
London EC2R 8AH
Contents

1 Concepts of statistical quality

1.1 Definition of Quality 4
1.2 International Context 4
1.3 Implications of differences between the ONS and the Bank 8

2 Data quality dimensions

2.1 Relevance 13
2.2 Accuracy 14
2.3 Timeliness and punctuality 18
2.4 Accessibility and clarity 19
2.5 Comparability 20
2.6 Coherence 21

Annexes
Annex 1: Statistical analysis of revisions data 24
Annex 2: Seasonal adjustment 25
Annex 3: Netting and off-setting 27
Annex 4: Confidentiality and Disclosure Control Policy 28
Annex 5: Statement of Administrative Data Sources 29

Abbreviations and Glossary 30

References 32
1 Concepts of statistical quality

1.1 Definition of quality

1.1.1 Quality is one of the key principles set out in the Bank’s Statistical Code of Practice,(1) which is referred to as the Code. In the Code, quality is defined as the fitness for purpose of published data for users and it encompasses a broad range of criteria. The Code requires that:

Quality standards for statistics will be monitored, and reliability indicators for individual data series will be progressively developed and published to assist users.
- Indicators of quality are being progressively developed for key statistical outputs.
- Statistical publications and publicly accessible databases will indicate where information on data quality may be found.
(Code of Practice, Section 3.2.2)

1.1.2 This definition is similar to that adopted by the UK Office for National Statistics (ONS), set out in the ONS Guidelines for Measuring Statistical Output Quality:(2)

The quality of statistical outputs is most usefully defined in terms of how well outputs meet user needs, or whether they are ‘fit for purpose’. This definition is a relative one, allowing for various perspectives on what constitutes quality, depending on the intended uses of the outputs.
(ONS Guidelines for Measuring Statistical Output Quality, Section A3)

1.1.3 The Bank collects a range of administrative data(3) (including regulatory data) as well as statistical data. The Bank also reserves the right to use collected statistical data for non-statistical purposes, as explained in Appendix 1 of the Statistical Code of Practice.

1.1.4 The new dual responsibility of the Statistics Division for both statistical and regulatory data collections opens up possibilities for using data more widely and efficiently across the Bank’s several functions: for monetary policy, financial stability and prudential supervision of financial institutions. It may also be possible to rationalise multiple collections from the same institutions, improving cost efficiency and reducing regulatory burden. These issues are being systematically reviewed.

1.2 International context

1.2.1 The measurement and reporting of data quality for official statistics are encouraged by international bodies. 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

---

www.bankofengland.co.uk/statistics/pages/about/code.aspx.
(3) Information on the Bank’s Statement of Administrative Data Sources can be viewed in Annex 5 of this document.
of the DQAF dates from 2012, and covers both generic and domain-specific statistical standards.\textsuperscript{(1)} 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 above. The United Kingdom 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 the ONS and by the Bank. An SDDS subscriber has to submit information about its data and its dissemination practices — 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 \textit{Dissemination Standards Bulletin Board}.\textsuperscript{(2)}

1.2.2 An important international initiative concerning the production of official statistics has been the development of the Generic Statistical Business Process Model (GSBPM), which is sponsored by the United Nations Economic Commission for Europe (UNECE) under its statistical metadata initiative, known as METIS.\textsuperscript{(3)} The GSBPM attempts a comprehensive itemisation of all stages in the production of good official statistics from, for example, determining users’ information needs to imputation, validation and archiving of output releases, and to the post-evaluation action plan. In all there are 47 of these stages under nine headings, as set out in \textit{Figure 1} on page 6.\textsuperscript{(4)} But the GSBPM is not a rigid framework in which all steps must be followed in a strict order; rather, it is an elaboration of all the possible steps and the interdependencies between them. The GSBPM should therefore be applied and interpreted flexibly. In the United Kingdom, the GSBPM is being developed by the ONS as a successor to its earlier Statistical Value Chain concept.

1.2.3 The Statistical Data and Metadata Exchange format, known as SDMX,\textsuperscript{(5)} is an international initiative sponsored by the Bank for International Settlements, the European Central Bank (ECB), Eurostat, the IMF, the OECD, the UN and the World Bank, which commenced in 2002, and is beginning to gain traction as it is adopted by institutions. The objective of SDMX is to ‘foster standards for the exchange of statistical information’ through the specification of common technical standards and IT architectures.

1.2.4 Within the European Union\textsuperscript{(6)} the relevant starting point is the set of standards set out in the Eurostat Data Quality Definition, adopted in 2003 following study and development lasting almost a decade. The Data Quality Definition became part of the European Statistics Code of Practice applicable to the European Statistical System (ESS), comprising Eurostat and the National Statistical Institutes (NSIs) of the Member States, under Recommendation COM(2005) 217 of the European Commission in 2005.\textsuperscript{(7)} The European Statistics Code of Practice covers similar ground to the IMF DQAF, comprising high-level principles covering ‘institutional environment’ issues, such as good governance standards, as well as more specific standards on the quality of statistical outputs taken from the ESS definitions. These definitions are summarised in \textit{Table A}. The Code was revised in 2013\textsuperscript{(8)} and provides a set of indicators of good practice for each principle.

\textsuperscript{(3)} UNECE (2009), \textit{Statistical Metadata (METIS)}, www.unece.org/stats/metis.html.
\textsuperscript{(5)} Statistical Data and Metadata Exchange, http://sdmx.org/.
\textsuperscript{(6)} The quality criteria for European Statistics are defined in European Statistical Law, Regulation (EC) 223/2009, Article 12 (at the time of publication this regulation was under revision). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:en:PDF.
Figure 1
Generic Statistical Business Process Model (GSBPM)

1. Simplify needs
   1.1 Determine needs for information
   1.2 Consult and confirm needs
   1.3 Establish output objectives
   1.4 Identify concepts
   1.5 Check data availability
   1.6 Prepare business case

2. Design
   2.1 Design outputs
   2.2 Design variable descriptions
   2.3 Design data collection methodology
   2.4 Design frame and sample methodology
   2.5 Design statistical processing methodology
   2.6 Design production systems and workflow

3. Build
   3.1 Build data collection instrument
   3.2 Build or enhance process components
   3.3 Configure workflows
   3.4 Test production system
   3.5 Test statistical business process
   3.6 Finalise production system

4. Collect
   4.1 Select sample
   4.2 Set up collection
   4.3 Run collection
   4.4 Finalise collection

5. Process
   5.1 Integrate data
   5.2 Classify and code
   5.3 Review, validate and edit
   5.4 Impute
   5.5 Derive new variables and statistical units
   5.6 Calculate weights
   5.7 Calculate aggregates
   5.8 Finalise data files

6. Analyse
   6.1 Prepare draft outputs
   6.2 Validate outputs
   6.3 Scrutinise and explain
   6.4 Apply disclosure control
   6.5 Finalise outputs

7. Disseminate
   7.1 Update output systems
   7.2 Produce dissemination products
   7.3 Manage release of dissemination products
   7.4 Promote dissemination products
   7.5 Manage user support

8. Archive
   8.1 Define archive rules
   8.2 Manage archive repository
   8.3 Preserve data and associated metadata
   8.4 Dispose of data and associated metadata

9. Evaluate
   9.1 Gather evaluation inputs
   9.2 Conduct evaluation
   9.3 Agree action plan

### Table A

**European Statistical System (ESS) definitions of quality dimensions**

<table>
<thead>
<tr>
<th>ESS quality dimension</th>
<th>ESS description (abbreviated)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relevance</strong></td>
<td>Relevance is the degree to which statistics meet current and potential users’ needs. It refers to whether all statistics that are needed are produced, the extent to which concepts used (definitions, classifications etc) reflect user needs, and the extent to which users are satisfied with the disseminated statistics.</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>Accuracy in the general statistical sense denotes the closeness of computations or estimates to the exact or true values. Statistics must accurately and reliably portray reality.</td>
</tr>
<tr>
<td><strong>Timeliness and punctuality</strong></td>
<td>Timeliness of information reflects the length of time between the availability of data and the event or phenomenon they describe.</td>
</tr>
</tbody>
</table>
| **Accessibility and clarity** | Accessibility refers to the physical conditions in which users can obtain data: where to go, how to order, delivery time, clear pricing policy, convenient marketing conditions (copyright, etc), availability of micro or macro data, formats, etc.  
Clarity refers to the data’s information environment: whether data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality is also available (including limitations in use) and the extent to which additional assistance is provided. |
| **Comparability**     | Comparability aims at measuring the impact of differences in applied statistical concepts, definitions, and measurement tools/procedures when statistics are compared between geographical areas, non-geographical domains, or over time. |
| **Coherence**         | Coherence of statistics is their adequacy to be reliably combined in different ways and for various uses. It is, however, generally easier to show cases of incoherence than to prove coherence.  
When originating from a single source, statistics are normally coherent in the sense that elementary results derived from the concerned survey can be reliably combined in numerous ways to produce more complex results. When originating from different sources, and in particular from statistical surveys of different nature and/or frequencies, statistics may not be completely coherent in the sense that they may be based on different approaches, classifications and methodological standards. |

1.2.5 Similarly, the 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 United Kingdom 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.

1.2.6 The importance of statistical quality to policymakers and other users is emphasised, for example, in the Ecofin Council conclusions of November 2012, particularly in the context of statistics underlying the Excessive Deficits Procedure and the Macroeconomic Imbalances Procedure.

1.2.7 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 United Kingdom by the ONS as the basis for data quality measurement and reporting set out in their ‘Guidelines’ framework. That framework comprises a core set of eleven key quality measures and a larger number of other quality measures, quality indicators and guidelines. The key quality measures are prioritised as the minimal quality data set for all ONS statistical outputs where they are relevant.

1.2.8 As discussed below, in a number of respects the more specialised characteristics of central bank data collections mean that not every ONS quality measure will be appropriate to the Bank’s statistical outputs. But where they are appropriate, there is merit in adopting international data quality standards that are widely recognised.

1.3 Implications of differences between the ONS and the Bank

1.3.1 Various differences arise in the data collection methods employed by national statistical institutes responsible for broader macroeconomic data, such as the ONS, and by central banks responsible for monetary and financial data, such as the Bank. These differences, set out schematically in Figure 2, affect the specification of appropriate statistical quality measures in the two frameworks.

1.3.2 Macroeconomic data for business and household sector data items typically depend on surveys, normally conducted using random sample or stratified random sample techniques. Other than for large businesses, sampling ratios will usually be small, giving rise to a mix of sampling and non-sampling errors. The extent of sampling errors can be controlled and estimated using statistical sampling theory, and will be a function of sample size. Sources of non-sampling errors can be mitigated through good survey design and practices such as data cleansing, imputation and other techniques, but will be independent of sample size. In random sampling frameworks, the estimated sampling error can be a data quality measure of accuracy.

Figure 2
Schematic data collection and sources of error: sampling and 'cut off the tail' frameworks

Sampling framework (eg ONS)
Population estimates
Estimated from sample totals. Introduces sampling and non-sampling errors.

Sample
Randomly drawn from population. Higher sampling ratio, up to 100%, for larger reporters under stratified sampling.

Response
Imputation used for non-responses. Introduces non-response bias and increases sampling errors.

Individual reporters
Data reporting errors detected through data cleansing. Incorrect data may be corrected (edited) at reporter level.

'Cut off the tail' framework (eg Bank of England)
Population measures
Aggregate of reporters' data and imputed data for tail entities. Introduces non-sampling error.

Reporting panel
Non-random panel determined by statisticians. Moderate coverage in terms of numbers of reporters; high coverage, approaching 100%, in terms of data quantities to be measured.

Response
Approximately 100% of reporting panel. Negligible non-response error.

Individual reporters
Data reporting errors detected through data cleansing. Reporters submit revised data or data adjustments are applied to output totals.
issuers of debt securities and arrangers or managers of such issues).(1) 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.

1.3.4 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.

‘Cut off the tail’

1.3.5 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.(2)

1.3.6 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 particular 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.

1.3.7 Under ‘cut off the tail’ sampling, there can be no formal estimation of sampling errors unless assumptions are made as to the properties of the tail reporters. However, with very high coverage it might be reasonably concluded that accuracy errors become correspondingly small. For this reason, coverage measures can be suitable quality indicators for central bank data.

Additional differences

1.3.8 Statistical reporting by banks and building societies can sustain a much greater complexity of data reporting than it would be reasonable for the 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.

1.3.9 There are certain differences in approach in the treatment of identified or suspected data reporting errors revealed by data cleansing. The ONS 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 (i.e., replacing with an imputed value) the data stored in their internal database system. The 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.

1.3.10 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 key quality measures, application of the adjustments facility in the Bank’s system cannot be interpreted as a pure indicator of data errors.

1.3.11 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.(1) 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.

1.3.12 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.(2) A high coverage of institutions is required to ensure that all banks above the CRD threshold are included.

1.3.13 Another distinction relates to the reliability of data inputs reported to the central bank. Given the heritage of UK banks’ statistical reporting that has evolved over the half century since the Radcliffe Report,(3) including frequent two-way communication between reporters (individually and collectively) and the Bank’s statistical staff, there tends to be 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.

1.3.14 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 the 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.

---

(1) One exception is Divisia money, which is stated at current prices and in indexed form. See ‘A new measure of Divisia money’, by Matthew Hancock, Bankstats, January 2005, www.bankofengland.co.uk/statistics/Documents/ms/articles/art2jan05.pdf.
(2) Cash ratio deposits are provided for under the Bank of England Act 1998, Schedule 2.
1.3.15 At present, little regulatory data are published as statistical outputs. Firms report using a common suite of regulatory returns, but the primary purpose is to support the regulator’s analysis of a given firm. The Bank publishes statistical data based on regulatory data submitted by some firms: jointly with the Financial Conduct Authority (FCA) the *Mortgage Lenders and Administrators statistics* are published quarterly\(^{(1)}\) and from 2013 the Bank publishes credit union statistics\(^{(2)}\). This list may be augmented in the future, for example, if aggregates are required from regulatory data sources to support the Bank’s analytical needs for macroprudential policy. In addition the Bank foresees requirements on firms to publish their own regulatory data, except where there is a good reason not to. This could lead to greater availability of published aggregated data, either from third parties or the Bank.


2 Data quality dimensions

This section describes in more detail the availability of data quality information, including definitions of specific quantitative measures, for the Bank’s statistical data. It is arranged by separate subsections according to each ESS dimension of data quality.

For the most part, the approach of this Framework is to summarise and refer to existing information on statistical data quality previously published by the Bank, where this is appropriate to the particular data quality dimension. In addition there are new, or formalised, definitions of quantitative measures of data accuracy, relating to coverage and revisions.

2.1 Relevance

2.1.1 The Bank’s Statistical Code of Practice discusses the implications of relevance in some detail. The top-level principle is that:

‘Statistics prepared by the Bank of England will inform significant decisions in the Bank, Government, business and the wider community. They also enable the Bank to be held to account.’
(Code of Practice, Section 1)

2.1.2 Relevance covers a range of issues such as:
• awareness of key users of the data and their needs;
• the relevance of the data to the needs of users;
• conformity of the data with relevant international standards; and
• information relating to the source of the data, including constraints on its availability.

2.1.3 The production of monetary and financial statistics is directly relevant to the Bank’s monetary policy responsibilities. Identification of ‘key users’ includes not only the Monetary Policy Committee (MPC) and the Bank’s analytical staff, but also external users of the data in the FCA and ONS. The ONS is a user of certain financial statistics outputs relating to the Monetary Financial Institutions (MFI) sector in the National Accounts, which in turn are used by the Bank in the conduct of monetary policy analysis. In addition, there is a broad range of users defined by the general public interest in economic statistics used for monetary policy analysis, and by the Statistics Division’s commitments to supply statistical data to international organisations, including the ECB, IMF and OECD.

2.1.4 Bank statistical data also contribute to financial stability analysis, both internally and externally. For example, data are supplied to the Bank for International Settlements (BIS) so that it can compile data on cross-border banking exposures, and the Bank co-operates with the BIS in its triennial surveys of foreign exchange and derivatives markets and its semi-annual OTC Derivative Survey.
2.1.5 With the recent establishment of the Bank’s Financial Policy Committee and the Bank’s new statutory responsibility for financial stability,(1) there is increased demand for statistical and non-statistical data of relevance to financial stability. Internationally, there is increasing demand for data relevant to systemic risks from the Financial Stability Board (FSB) and the European Systemic Risk Board (ESRB). The development of statistics for financial stability is in its infancy, and presents a fresh set of challenges, including devising and implementing appropriate quality standards.

Available information

2.1.6 The Bank generally uses articles in Bankstats(2) or in the Quarterly Bulletin(3) to consult and update users of statistics as required under its Statistical Code of Practice.(4) The Code makes four specific commitments in Section 1:

- Producers of statistics within the Bank have a special relationship with key users in the Bank, and with the FCA and the ONS, but will manage these relationships in ways that minimise any disadvantage to other users.
- The relevance of existing statistics to the needs of users will be monitored and material gaps in their coverage identified.
- The views of users will be sought when important changes to statistical collections, methods of compilation, or outputs are planned.
- Users will be kept informed about the Bank’s statistical development work programme and will be encouraged to influence its future direction.(5)

2.2 Accuracy

2.2.1 Accuracy means the closeness of the estimated values to the (unknown) true values. In practical terms, for economic statistics accuracy means the evaluation of potential sources of error.(6) Generally speaking, sources of inaccuracy can arise from missing data on account of non-respondents, use of imputation, incomplete coverage in random sampling or ‘cut off the tail’ samples, or from human error. For given sample size and response rates, methods of minimising data errors include maintaining clear statistical reporting guidance, liaising with reporting institutions and data cleansing.

2.2.2 For Bank data, response rates can be regarded as effectively 100%. It is possible, however, that reporters might submit data known to be incorrect or which are not compliant with statistical requirements. For example, a reporter might submit estimated data based on previous period outcomes, or data on an accounting basis that departs from the required statistical reporting basis. This might be with the Statistics Division’s understanding (a practice loosely termed ‘best endeavours’ or better, ‘agreed acceptable endeavours’ reporting) or otherwise. In recent years there has been a shift towards more specific statistical reporting guidelines and the scope of ‘endeavours’ reporting has been reduced.

2.2.3 The quality of source data submitted by reporters is continuously monitored through processes of data cleansing. These consist of data validation, an automatic process in which all

---

(1) This was done through the Financial Services Act 2012, which can be viewed here: www.legislation.gov.uk/ukpga/2012/21/contents/enacted.
(2) Bankstats is a collection of pre-defined tables providing more detailed breakdowns and longer time series of data published in the Bank’s statistical releases, as well as other statistics compiled by the Bank. See www.bankofengland.co.uk/statistics/Pages/bankstats/default.aspx.
(5) An annual statement of the Statistics Division’s statistical work programme, eg Bailey and Owadi (2013), is published on the Bank’s website. This covers planned improvements to statistical outputs, including those related to ONS priorities. Available at www.bankofengland.co.uk/statistics/Documents/ms/articles/art1apr13.pdf.
relevant in-form and cross-form accounting identities are encoded into the reporting forms and are checked on receipt, and plausibility checking. Plausibility checking involves raising questions of reporting institutions when possibly abnormal data responses are detected, such as data inputs that lie outside of an expected range of values, or which otherwise raise concerns. Where appropriate, analysts in the Statistics Division will enquire if reported data are consistent with news of any known financial market developments which have significant quantitative implications, such as securitisations, or mergers and takeovers.

2.2.4 Measures of coverage and of revisions represent two key quantitative dimensions relating to accuracy. These are considered in turn.

Coverage

2.2.5 Coverage measures how much of the total population of reporters is included in the reporting sample. As reporting institutions vary considerably in size, an appropriate quantitative indicator should take account of the aggregate size of the reporting institutions in that of the total population. This requires a suitable choice of size measure.

Definition: coverage

The coverage of a statistical output can be defined with respect to the reporting panel of the statistical return required for its derivation. It will normally be calculated as the aggregated balance sheet of the reporting panel expressed as a percentage of the aggregated balance sheet of the total population of reporters for the most recent end-calendar year observation. Where appropriate, coverage can instead be calculated in terms of an alternative size measure more relevant to the data collected by the statistical return.

2.2.6 This definition means that any family of outputs derived from the same statistical reporting form will have the same coverage, and therefore it is a measure that need only be evaluated on a form-by-form basis. It is a well-defined measure if a 100% survey of balance sheet size (or of any alternative size measure) is maintained for at least an annual frequency of reporting. It should remain practically well-defined if very small institutions are relieved from reporting requirements, so long as reliable estimates indicate that their contribution to the aggregate size measure of reporting institutions would be immaterial.

2.2.7 Scaling by balance sheet size is suggested as the default measure. The alternative coverage definition, however, will enable those cases of statistical outputs where calculations of coverage would be unlikely to be well proxied by balance sheet size to be sensibly measured. A number of the statistical reporting forms other than the core balance sheet return, form BT, are designed to gain more detailed information on particular subsets of the balance sheet, for example, external assets and liabilities. In such cases, appropriate alternative coverage measures might be based on the relevant subset of the balance sheet or on a relevant form BT box. Any alternative measure should be based on quantities for which there are census estimates or for which overall coverage can be reliably assumed to be close to 100%.

2.2.8 Any suitable indicator of coverage must have the property that it increases with the inclusion of additional reporters. Choices of coverage measure should therefore be defined by reference to quantities reported with a consistent arithmetic sign. Definitions based on netted quantities, for which positive and negative outcomes might each be reported, will not serve as suitable coverage measures because of the potential for contributions from different reporters to cancel in aggregate.
**Imputation**

2.2.9 Imputation is used to estimate unavailable data items for specific reporting institutions on particular statistical returns. Data may be unavailable because they relate to institutions not required to report at the highest frequency or to those not in the reporting panels for the statistical return, see paragraphs 1.3.5–1.3.7 on ‘cut off the tail’. The use of imputation enables aggregate measures to be estimated from the data available from reporting panels that are considerably smaller than the full population, reducing the reporting burden on smaller or less relevant institutions. There are various examples of imputation methods.\(^1\) The simplest methods involve interpolation of data between reporting periods, or projection of data beyond last available periods, in order to generate higher-frequency estimates with respect to non-reporters for intermediate and most recent periods. Other approaches entail application of average characteristics of the reporting panel population, or presumed accounting and behavioural relationships, to generate estimates for non-reporters.

2.2.10 The impact of imputation on accuracy will depend on the relative contribution of the imputed data to any estimated data output. This can be viewed as the complement of the coverage of the data estimate, where coverage is defined as the amount of the estimate explained by actual inputs from reporters. If coverage is high, as is generally the case, the impact of any inaccuracies resulting from imputation assumptions should be small.

**Revaluation adjustments**

2.2.11 The Statistics Division follows international guidance on the estimation of data on financial flows, or transactions in financial assets and liabilities, as set out in the IMF Monetary and Financial Statistics Manual.\(^2\) Estimates of financial flows are derived from changes in levels of financial assets and liabilities during the period, less adjustments made for estimated revaluations and for ‘other changes in the volume of assets’ (OCVA). Revaluations of foreign currency denominated assets and liabilities are computed from movements in end of period exchange rates and data on the currency detail of reporters’ balance sheets. Reporting guidelines require specified balance sheet items to be reported according to end-period exchange rates and market prices, and securities held as assets should be reported at market value. Revaluation effects also occur and are taken account of with respect to equity price and other financial asset price changes.

2.2.12 OCVA encompasses all examples of changes in assets and liabilities not due to financial transactions or revaluation effects. For monetary and financial data, the most important component of OCVA is debt write-offs. The statistical reporting form WO, introduced in 2008, was designed to improve the reporting of write-offs data. Form WO entailed a higher frequency and more detailed analysis of write-offs data than was previously available.

**Revisions**

2.2.13 Revisions are a measure of how close initial estimates are to later or settled estimates of the data. They relate to the reliability of the initial estimates as indicators of the final estimates.

2.2.14 Revisions can result from the receipt of new data not available for earlier published estimates, including late and corrected submissions of data by reporters following the identification of errors in the original data. They may also arise, however, from the application of methodological improvements such as better imputation rules, definitions moved to a sounder theoretical basis, or from other adjustments to the back data. It follows that evidence of

---


revisions need not be interpreted solely as an adverse indicator of data quality. In the case of seasonally adjusted data, new data observations will generally cause past seasonal factors to be updated, so that back data are routinely revised even when the unadjusted back data are unchanged.

2.2.15 The incidence of data revisions will depend on the revisions policy applied by the statistical authority and has to be interpreted in context. Section 3.4 of the Bank’s Statistical Code of Practice states:

Data revisions will be transparent and will be governed by a clear policy applied consistently across all outputs.

(Code of Practice, Section 3.4)

2.2.16 Under the Code statistical outputs remain indefinitely open for revision, and revisions are required to be applied at the earliest available reporting date. The Statistics Division will normally pre-announce and where appropriate explain in Bankstats the application of methodological changes and any new imputation assumptions resulting in revisions to past data.

2.2.17 Institutions reporting to the Bank can, and do, revise their data submissions at any time in the reporting cycle, including in the period before the first estimate of the reporting period is released. Analysis of data input revisions from the monthly balance sheet return, form BT, and the quarterly income and expenditure return, form PL, reported by Burnett and Foster in Bankstats in 2007 revealed that by far the greatest number of data revisions by reporters occurred before the first release, and that very few revisions were made to data more than three to four periods ago.\(^{(1)}\) They concluded that ‘only in exceptional circumstances would data [at any frequency] not be [considered] final at the two-year horizon’. Two years is therefore used as the de facto horizon by which data can be assumed to have stabilised in the definitions to be considered here.

Definition: revisions

Two quantitative indicators of revisions can be defined: the mean revision and mean absolute revision, where revisions are calculated as the change from the first published estimate to the estimate published after a fixed interval of two years. Mean and mean absolute revisions should be calculated over a recent sample of, if possible, at least three whole calendar years.

Where the data are available and they make a material contribution to the size of revisions, an analysis of revisions distinguishing the contribution of methodological improvements may be provided.

2.2.18 This definition is similar to that adopted in the ONS Guidelines for Measuring Statistical Quality. The mean revision can be interpreted as the average bias in original estimates. The mean absolute revision is a measure of dispersion of revisions. Standardisation of the definition in terms of the revisions interval and calculation period enables users to have confidence that the measure is objectively defined. The choice of the two-year fixed interval followed the Burnett and Foster (2007) analysis and Statistical Code of Practice guidance on how quickly series tend to stabilise at their final values.

2.2.19 The definition does not prescribe particular forms of presentation of revisions data, for example whether revisions should be expressed in levels, flows or growth rate quantities. Often, data on revisions are presented as scaled measures eg mean absolute revisions divided by mean level or by a measure of the variance of the underlying series. The choice as to which forms of presentation are most informative will depend on the statistical properties of the series in question. A number of detailed recommendations on the presentation of summary data for revisions were discussed in an OECD/Eurostat taskforce (2007).\(^{(1)}\)

2.2.20 Assessing the statistical significance of any revisions, such as whether they indicate evidence of bias in the original estimates, is discussed in Annex 1.

2.2.21 Analysis of revisions to Bank monthly monetary data series were first reported by Franklin in Bankstats in July 2005.\(^{(2)}\) This analysis covered ten key monetary series produced on a monthly frequency for reporting periods from January 2000 to July 2004. The revisions between the one-month and ten-month estimates were examined for each series, and corresponding bias, standard error and 95% confidence intervals derived. The standardised definition in this Framework adopts a longer fixed interval than the one used in 2005 in order to be more consistent with the Statistical Code of Practice and the later evidence on the stability of statistical outputs after two years.

**Available information**

- The Statistics Division’s approach to data cleansing including plausibility checking was described by Bigwood in Bankstats in 2004.\(^{(3)}\)
- The Statistics Division discloses the full set of current reporting panel criteria on the Bank website.\(^{(4)}\)
- Statistical releases identify material revisions, and the thresholds for materiality, as required under the Bank’s Statistical Code of Practice.
- The Statistics Division regularly updates its analysis of revisions.\(^{(5)}\)

### 2.3 Timeliness and punctuality

2.3.1 Timeliness means how soon the scheduled publication date comes after the reference period. Release schedules are pre-announced and are available under the Statistical Interactive Database.\(^{(6)}\) Other things being equal, the sooner the data are published the more valuable for users, and in this sense timeliness is a simple indicator of data quality. However, there is a trade-off between the benefits of greater timeliness for users and of more time for reporters and compilers to ensure fuller coverage and more thorough plausibility checking. The Bank’s Statistical Code of Practice requires a balance between benefits to users and the costs to reporters:

Statistics will be made available as early as possible after they have been compiled, either by formal release or by announcing that they are available on request.

(Code of Practice, Section 3.3.1)

---


2.3.2 The Bank’s approach to timeliness and flexibility follows that to be found in the IMF’s Special Data Dissemination Standard 2013: Guide for Subscribers and Users. (1)

2.3.3 Punctuality relates to the extent to which data releases are published on time according to the pre-announced publication dates, and is monitored closely. In practice, the Bank’s statistical data are routinely released on time.

*Available information*

- The statistics releases schedule is pre-announced, and can be viewed via the Statistical Release Calendar. (2) Individual release dates are also available through selecting the relevant subject.
- Timeliness can be measured by the time lag in days between the end of the reference period and the release date of the first estimate. The Statistics Division will periodically review the timeliness of its statistical outputs against international practices.

### 2.4 Accessibility and clarity

2.4.1 Accessibility and clarity concern the ease with which users can access the data, including access to interpretation and guidance. Developments to enhance accessibility of the Bank’s monetary and financial statistics data have included the launch of the Statistical Interactive Database and the introduction of the Data Updates Facility, which provides users with a history of data revisions for any data observation.

2.4.2 The Statistics Division makes data quality information available through two principal channels: through periodic articles in *Bankstats*, and through the Explanatory Notes (or metadata) attached to statistical releases and to all data series in the Bank’s Statistical Interactive Database. Through a project undertaken in 2007, the Bank adopted a consistent format for metadata content under common headings. (3) These headings cover:

- Overview
- Definitions
- Availability
- Valuation and break
- Sources
- Further information

2.4.3 Consistently with wider adoption of electronic media, monthly *Bankstats* has been published in online format only, since July 2006. Online availability was first introduced in 2004, with hard-copy publication limited to two editions per year. Final phasing out of hard-copy publication reflected user feedback favouring online publication and cost efficiency considerations. Reporting forms, data definitions, and statistical bulletins addressed to reporting institutions are also available online, allowing users as well as data providers to be well-informed on data definitions. *Bankstats* is now a monthly online publication with fixed monthly release dates. The most current data releases can be accessed via the Interactive Database, updated at 9.30 am daily.

*Available information*

- Substantial metadata information under standardised headings is made available on the Statistical Interactive Database and in statistical releases. *Bankstats* articles convey further technical information.

---

2.5 Comparability

2.5.1 Comparability concerns the degree to which data may be compared over time or across domains.

2.5.2 Section 3.1 of the Bank’s Statistical Code of Practice promotes the adoption of internationally accepted standards, guidelines and norms of good practice. Bank of England monetary and financial statistics data should conform to the European System of Accounts (ESA 1995, ESA 2010 from September 2014); and also to the IMF Balance of Payments Manual (BPM 6, also from September 2014) and Monetary and Financial Statistics Manual (MFSM), the more specialised subsidiary manuals to the UN 2008 System of National Accounts (SNA 2008). The ESA standard has statutory authority under European legislation, but is drafted with the express objective of maintaining consistency with the SNA. Differences between ESA and SNA are mainly confined to ESA’s provision of appropriate statistical definitions for euro-area measures, to certain instances where ESA is more prescriptive than the equivalent SNA standards, and to some differences in terminology. For most practical purposes, however, these standards can be considered to be consistent with each other.

2.5.3 General notes and definitions for reporting institutions, detailed guidance on data definitions and validation checks specific to each reporting form and updated guidance in the form of regular statistical notices are maintained and made available to reporting institutions. This compendium is also known historically as the Yellow Folder, although it now exists only online. Under the Bank’s Statistical Code of Practice, financial reporting standards are normally accepted for statistical reporting so long as they are considered to be consistent with international statistical standards, or differences from these standards are not judged to be material.

2.5.4 Information on comparability is provided as part of the Statistical Interactive Database metadata. Changes to the definition or coverage of statistical outputs are pre-announced (and where appropriate open to prior consultation) and explained in Bankstats or the Quarterly Bulletin. Examples of information on comparability include:

- The full set of statistical definitions and standards in use, as set out in the Yellow Folder.
- Conventions on break adjustments and the calculation of growth rates.
- Routine announcements of revised definitions.
- Routine announcement of any ONS classification changes affecting banking counterparties.

Seasonal adjustment

2.5.5 Comparability of data over time is enhanced through the application of seasonal adjustment methods. The objective of seasonal adjustment is to produce adjusted time series free of significant calendar or seasonal effects. As a matter of policy, the Bank commits to publishing non seasonally adjusted versions of any seasonally adjusted data series that it publishes. However, there is no presumption that any particular data series should be subject to seasonal adjustment. Further discussion of the principles of seasonal adjustment can be found in Annex 2.

2.5.6 Seasonal adjustment of Bank statistical data is currently undertaken using the X-12-ARIMA package, which replaced the GLAS (generalised linear abstraction of seasonality) package in 2004. Use of the X-12-ARIMA package and knowledge of the history of the series allow appropriate parameter settings, including a RegARIMA model (that may include calendar

and other regressors) and length of trend and seasonal filters, to be determined. The quality and stability of the seasonal adjustment of any series can be assessed by reference to diagnostics including standard X-12-ARIMA diagnostics.

2.5.7 When X-12-ARIMA was adopted, the division undertook to review regularly around 300 individual series to determine whether seasonality remained present and whether any modification of the seasonal adjustment parameters was required. Within this total, attention is focused on those which are key published series, or for which preliminary analysis suggests scope for improving the adjustment, eg the need to review current regressors. Where seasonal effects are no longer considered significant the seasonally adjusted series is discontinued. Summary reports of these regular reviews and an annual report updating users have been reported in Bankstats since 2005. Significant changes to seasonal adjustment practices (for example, the adoption of X-12-ARIMA in 2004) are also consulted upon in Bankstats. The ONS has a guide(1) to seasonal adjustments for time-series analysis.

2.5.8 The Statistics Division has announced that it is planning to switch over to the X-13ARIMA-SEATS seasonal adjustment package, most likely at some point during in 2014, following an evaluation of this package conducted by the Government Statistical Service in which the Bank and the ONS also participated.(2) X-13ARIMA-SEATS allows for much of the methodology of X-12-ARIMA to be retained alongside additional new methods, and is maintained as the official successor to X-12-ARIMA by the US Census Bureau, who are also responsible for its development and distribution. It therefore commands wide international acceptance.

Available information
• The Statistics Division conducts annual reviews of seasonal adjustment and reports the outcomes.

2.6 Coherence

2.6.1 Coherence relates to the degree to which data derived from different sources or methods but concerning the same phenomenon are consistent with each other. The accounting statement format of the main statistical reporting forms means that a high degree of coherence can be expected and internal consistency of monetary and financial data sets can be readily tested. At the lowest level, validation checks impose consistency across data reported by individual banks in their statistical returns. At higher levels, assessing consistency across aggregated totals may reveal the effects of any sectorisation, classification and other errors undermining data coherence. Examples of investigations which promote coherence include:

• Cross-form plausibility checks on interest receivable or payable applied to the effective interest rates return, form ER, and the profit and loss return, form PL.
• Reconciliation of MFIs’ loan transfers data with those reported by Specialist Mortgage Lenders and with market intelligence.
• Checks against commercial data sources for quoted interest rates and capital issues data.
• Reconciliation of the inter-MFI difference: ie consistency of the aggregate measures of MFIs’ assets and liability positions with respect to each other.
• Reconciliation of sectoral and industrial analyses of lending and deposits.
• Comparison of UK Debt Management Office data with MFIs’ government debt data.

2.6.2 Coherence issues may arise with respect to financial accounts prepared under national and international financial reporting standards: UK Generally Accepted Accounting Principles (GAAP) under the Accounting Standards Board, and International Financial Reporting Standards (IFRS) under the International Accounting Standards Board (IASB), applicable at group level to listed EU companies from 2005 onwards. The use of IFRS is being integrated with financial reporting for regulatory purposes under the 2013 Capital Requirements Directive (commonly known as CRD IV).(1)

2.6.3 The introduction of IFRS raised a number of technical issues for statistical reporting on which the Statistics Division provided specific guidance in statistical notices, and will continue to consult reporters whenever changes to the standards are proposed(2) by the IASB.

Available information

- Yellow Folder guidelines addressing statistical reporting issues arising under IFRS.
- ‘Frequently asked questions relating to IFRS’. (4)

---

Annexes

The following annexes relate to certain methodological issues that have been particularly relevant to users of the monetary and financial data sets produced by the Bank of England. They cover:

- Statistical analysis of revisions data.
- Seasonal adjustment.
- Netting and off-setting.
- Confidentiality and Disclosure Control Policy.
- Statement of Administrative Data Sources.
Annex 1: Statistical analysis of revisions data

1 Users of data may be concerned to understand whether the magnitude of revisions, regardless of sign, is material in comparison to the information conveyed by the reported data, and whether there are systematic biases in the direction of revisions.

2 The standard statistical test for the presence of bias in a sample of revisions data establishes whether the mean revision is significantly different from zero taking account of the size of the sample and the variance of revisions, under the assumption that the revisions are independently and normally distributed: this is the t-test. For economic times series it is appropriate first to test whether the revisions data are significantly auto-correlated, in which case a modified form of t-test is recommended. (Alternative approaches to such tests for bias are discussed by Jenkinson and Stuttard (2004) and Di Fonzo (2005)).

3 However, tests for bias might be considered to be relevant only when the magnitude of revisions in absolute terms is, in some sense, large compared to the underlying data. One approach could be to investigate whether the magnitude of the mean square revision, a measure encompassing the bias and variance of revisions to a series, has a material impact on the final data.

4 Assessing materiality of revisions in the sense described here is a judgemental question for users rather than a formally testable hypothesis. One approach is to calculate a noise to signal ratio, such as the ratio of the mean square revision to the variance of published data series. If this ratio is small then the implication is that the revisions can be considered immaterial compared to the changes in the underlying series. Conversely, large values of a noise to signal ratio could be interpreted as evidence that the first estimates are unreliable. But a large value for this ratio might also arise if revisions are small in absolute terms but the underlying data also exhibit low variance. In this situation it does not necessarily follow that the original estimates of the data series are unreliable, which suggests that use of a noise to signal ratio can be inconclusive.

---


Annex 2: Seasonal adjustment

1 Many economic data series are materially affected by regular seasonal fluctuations or calendar effects. It is a conventional practice, which the Bank also follows, to attempt to identify, estimate and remove such effects from time-series data, in order to allow the underlying picture to be more clearly evident. This process is known as seasonal adjustment.

2 Regular seasonal fluctuations are those movements which, on the basis of the past movements of the time series in question, can under normal circumstances be expected to recur with similar intensity in the same season each year. Calendar effects relate to effects caused by the number of working days or calendar days in the month, or the dates of particular festivals or similar occasions, such as Easter. Within the year the influence of such effects on a particular month can vary from year to year, but there exists a basis for quantifying and adjusting them out.

3 In seasonal adjustment, time series are typically regarded as the composition of three elements: a slowly varying trend, a seasonal component and an idiosyncratic term, meaning shorter-term perhaps volatile impacts. The objective of seasonal adjustment is to identify when such a decomposition can reliably be estimated, and in which case to remove the impact of the seasonal component. It is important to appreciate, therefore, that seasonal adjustment is not simply the process of smoothing a volatile series.

4 Economic statistics can reflect the impact not only of fundamental movements in the data but also of ‘break’ effects such as classification or definitional changes. Typically such changes might have a step change impact on the level of a particular quantity but be unrelated to the underlying rate of change of the economic concept to be measured. For calculations of published statistics of flows and growth rates quantities, the Statistics Division’s standard practice is to adjust for the effects of material known breaks in the periods in which they occur.(1)

5 Such break adjustments are also provided for in the application of seasonal adjustment: this ensures that when the seasonal adjustment process is applied to levels series (which is the norm), that it is undistorted by the impact of breaks. The break effects are then reapplied so that they remain evident in the seasonally adjusted levels data. The intended outcome is that seasonally adjusted and non seasonally adjusted versions of a data series should either reflect, or be adjusted for, identical impacts of material known breaks.(2)

6 The nature and strength of seasonal effects will differ across data series, and may also vary over time for a given series. Seasonal adjustment software packages therefore encompass a range of seasonal adjustment techniques, or particular techniques may allow for alternative choices of parameter settings, for example, those governing trend and seasonal filters. The outcomes of alternative methods and settings are evaluated according to multiple criteria, and there need not in principle be a uniquely identifiable ‘best’ adjustment, nor necessarily a categorical determination of the presence of seasonal effects in any particular case.

7 Seasonal adjustment can be problematic in certain circumstances, notably when a new data series is under consideration or one that has been subject to a significant re-definition or whose properties have changed markedly on account of underlying changes in the economy. A minimum of three years of observations are required to carry out a seasonal adjustment. However, the seasonal adjustment of short series may be prone to more uncertainty and greater

(1) In addition, for a very limited number of more high-profile statistics, break-adjusted variants of levels data series are also published.

(2) Technical discussion of break effects in time series and the application of seasonal adjustment can be found on the General Notes pages: www.bankofengland.co.uk/statistics/pages/aeb/General%20Notes/default.aspx.
volatility than when more observations are available. A recent M4⁴⁰ seasonal adjustment article(1) alluded to this issue by stating: ‘It is recognised that seasonal adjustment of series with less than five years of data are subject to greater volatility. These series will therefore be monitored closely over the next year’.

8 Another issue concerns the best practices to be followed when considering the seasonal adjustment of related data series, for example, series representing a total and its components. There is typically no guarantee that the series relationship will consistently hold when the series are independently seasonally adjusted; conversely, imposing this constraint may force a second-best form of adjustment in some situations. Such considerations indicate that, while seasonal adjustment of data series is an automated part of the statistical routine, there remains a requirement for the application of judgement in considering how this is undertaken. The Statistics Division undertakes an annual review of its seasonally adjusted statistical outputs, reporting a summary of the results in the form of a regular Bankstats article.(2)

---


(2) For example, the 2012 review reported in May 2013: www.bankofengland.co.uk/statistics/Documents/ms/articles/art1may13.pdf.
Annex 3: Netting and off-setting

1 One feature of the monetary and financial statistics data is the application of the ‘off-setting’ rules, also known as ‘netting’. The general rule applies to individual reporters and requires, other than under specified circumstances, that they should record the full amounts of asset and liability positions, for example, loan and deposits amounts, with respect to the same counterparties. That is, if a bank’s customer holds both a deposit balance and a loan balance, each of these positions should contribute fully to their total reported deposits and loans, rather than that a net position contributing only to one side of the balance sheet. This accords with the general principle under the IMF Monetary and Financial Statistics Manual, and under the SNA 2008, that assets and liabilities should be recorded on a gross basis.

2 There are some circumstances, however, under which ‘off-setting’ is permitted under the Bank’s reporting rules. These are set out in Section 6 of its General Notes and Definitions document. Off-setting is permitted only for balance sheet forms, so not for the effective rates and profit and loss data, under the following conditions: the reporter enjoys a legal right of set-off with the counterparty; the off-setting balances are denominated in the same currency; the debit and credit balances relate to the same customer, or to customers in the same company group, and the netted accounts should be managed and controlled on a net basis.

3 It might be considered that the exceptions permitted under these conditions represent a compromise between strict adherence to the IMF principle of no netting, and recognition of the existence of certain credit or banking products in which netting is a practical reality. For example, the rules should in principle accommodate ‘off-set mortgage’ products managed as a single integrated account, or in the case of corporate customers, multiple account liquidity arrangements that are managed subject to a net overall credit limit. It is important to note also, however, that the Bank’s off-setting rules are stated in language identical to that contained in the Cash Ratio Deposits (Eligible Liabilities) Statutory Instrument, and derive from an earlier position when the bank’s statistical reporting forms were fully aligned to the CRD definitions. As CRDs are now reported on a dedicated Eligible Liabilities form, this need no longer represent a binding tie.

4 Reporters are asked to consult the Bank if they intend to change their off-setting treatment of data reported for statistical purposes. The Statistics Division’s preference, on grounds of statistical data quality and in order to conform better to the IMF guidelines, is in general to discourage extensions to the practice of off-setting in banks’ statistical reporting.

---

(1) The General Notes and Definitions document can be accessed here: www.bankofengland.co.uk/statistics/Pages/reporters/default.aspx.
Annex 4: Confidentiality and Disclosure Control Policy

1 The Bank and the ONS follow similar policies for maintaining the confidentiality of individual institutions’ data through disclosure control. The Bank’s policy is described in the Statistical Code of Practice Section 5.3.1: *Individual respondents will not be identifiable within published statistics*, and in the published confidentiality statement.\(^{(1)}\)

2 Practice for ONS is described in the Code of Practice of Official Statistics, Principle 5.1: *Ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information*. The ONS publishes a guide\(^{(2)}\) to its disclosure policy.

3 The Code of Practice for Official Statistics is the subject of guidance issued on behalf of the Government Statistical Service (GSS). The Bank’s practice is quite specific and reflects the structure of the banking industry. In practice any statistic containing source data from fewer than three institutions is withheld from publication unless the institutions’ consents have been given. Similarly aggregates or sub-aggregates may be suppressed if a single institution’s data can be indirectly deduced.

4 The GSS guidance is more generic reflecting the differing structures of industries and sectors across the economy. The rules for each sector are confidential since it is deemed that knowledge of the values used to derive thresholds reduces the protection. The rules can be summarised as:
   • There must be at least \(n\) enterprise groups in a cell (or statistic)
   • The total of a cell minus the largest \(m\) reporting units must be greater than or equal to \(p\)% of the value of the largest reporting unit

If a cell does not meet these two criteria it may be suppressed.

5 For instance, in 2013 the Bank ceased publishing separately data for building societies\(^{(3)}\) once it was clear that this series would be disclosive of one particular institution’s data. The Bank has further since changed to only publishing data for total Monetary Financial Institutions, to continue to prevent the risk of disclosure.\(^{(4)}\)

---


Annex 5: Statement of Administrative Data Sources

1 In its Statistical Code of Practice published in June 2013, the Bank committed to publishing and maintaining a Statement of Administrative Data Sources(1) documenting: the sources used for statistical purposes; untapped sources recognised to have the potential for such use; the arrangements for monitoring the quality of the source data; and procedures to acknowledge the implications for statistics when changes to administrative sources are planned.

2 Statistics published by the Bank under the Code are based on two main sources — data gathered under statistical statutory powers (for monetary policy and cash ratio deposit purposes), and data extracted from administrative or management systems to produce statistical outputs as a by-product.

3 Administrative and management data will have been initially collected for a variety of operational purposes including prudential regulation, monitoring and analysis, policy development, analysis of the effect of policy changes, forecasting, resource allocation and answering Parliamentary Questions and Freedom of Information requests.

4 By using data already available within administrative or management systems to produce statistics, rather than collecting data afresh, it is possible to limit the overall burden placed on data providers and avoid the costs of mounting dedicated data collection exercises. Additionally information extracted from such systems may have the advantages of being timelier than statistical data, and with greater breadth of coverage.

5 The UK Statistics Authority actively encourages public bodies to exploit administrative and management sources for statistical purposes. However the Authority recognises that the statistical advantages of such arrangements can only be fully realised if statisticians have appropriate access to such systems; if statistical purposes are reflected in the design, management, and development of such systems; and if adequate safeguards are put in place to ensure the professional integrity of published statistics derived from them. The Authority’s main requirements are set out in the third Protocol attached to its Code of Practice for Official Statistics.(2) Although the Bank is not bound by the Authority’s requirements, its own Statistical Code of Practice reflects similar principles.

---

Abbreviations and Glossary

**ARIMA** — Autoregressive Integrated Moving Average. ARIMA models are a family of statistical models used for modelling and forecasting time-series data which can be stationarised by applying transformations such as differencing and logging. Seasonal ARIMA models are used to efficiently model seasonal time series.

**Bias** — Systematic deviation from the true value, as distinct from random errors which cancel each other out in the long run.

**BIS** — Bank for International Settlements.

**BPM** — Balance of Payments Manual.

**Confidence interval** — Interval estimate which has a known and controlled probability to contain the true value of the population parameter.

**CRD** — Cash Ratio Deposits: non-interest bearing deposits required to be placed by banks and building societies with the Bank of England.

**CRD IV** — The fourth Capital Requirements Directive, which is the EU implementation of the new Basel III accord.

**DQAF** — Data Quality Assessment Framework.

**DQF** — Data Quality Framework.

**ECB** — European Central Bank.

**ESA** — European System of Accounts.

**ESCB** — European System of Central Banks: comprises the ECB and the national central banks of the EU Member States.

**ESRB** — European Systemic Risk Board: the EU macroprudential body, established in 2011, sits with the ECB in Frankfurt. It has the power to give public recommendations to support financial stability within the EU.

**ESS** — European Statistical System.

**EU** — European Union.

**Eurostat** — EU Statistical Office.

**FCA** — Financial Conduct Authority.

**FSA** — Financial Services Authority, now defunct, replaced by the Financial Conduct Authority and the Prudential Regulation Authority in 2013.

**FSB** — Financial Stability Board.

**GAAP** — Generally Accepted Accounting Principles.

**GLAS** — Generalised linear abstraction of seasonality.

**GSBPM** — Generic Statistical Business Process Model.

**GSS** — Government Statistical Service.

**IASB** — International Accounting Standards Board.

**IFRS** — International Financial Reporting Standards.

**IMF** — International Monetary Fund: an international organisation of 185 member countries established to promote international monetary co-operation, exchange stability and orderly exchange arrangements.

**Imputation** — A procedure for entering a value for a specific data item where the response is missing or unusable.
MFI — Monetary Financial Institution: in the United Kingdom this is the central bank, other banks and building societies. In other European countries, building societies may not exist but credit unions have a greater prevalence.

MFSCG — Monetary and Financial Statistics Compilation Guide.


MLA/MLAR — Mortgage Lenders and Administrators: mortgage lenders and administrators are required to submit data on financial information, quantitative and qualitative information, provisions of qualitative mortgage information and annual fee tariff measures.

MPC — Monetary Policy Committee.

NSIs — National Statistical Institutes.

OCVA — Other changes in the value of assets.

OECD — Organisation for Economic Co-operation and Development.

ONS — Office for National Statistics.


Random error — An error caused by random, unpredictable fluctuations from a true value.

RegARIMA models — Statistical model in which some features of the time series are modelled with linear regression variables while the remaining features (those of the regression residuals, including trend, cycle and seasonal components) are modelled with a seasonal ARIMA model.

Sampling error — Difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a sample of values is observed.

SDDS — Special Data Dissemination Standard.

Seasonal adjustment — Method used to identify, estimate and remove seasonal fluctuations from time-series data.

SEATS — Signal Extraction in ARIMA Time Series.

SNA — System of National Accounts.

Statistical microdata — An observation data collected on an individual object.

Stratified sample — A sample selected from a population which has been stratified, part of the sample coming from each stratum.

Time series — A sequence of data points in time, measured at approximately equally spaced times.

UNECe — United Nations Economic Commission for Europe.

X-11 — Seasonal adjustment software developed by the US Census Bureau, based on an iterative application of linear filters.

X-12-ARIMA — A seasonal adjustment software package, used to estimate and remove regular seasonal fluctuations and typical calendar effects.

X-13ARIMA-SEATS — Seasonal adjustment software under development at the US Census Bureau in collaboration with the Bank of Spain that integrates an enhanced version of X-12-ARIMA with an enhanced version of SEATS to provide both X-11 method seasonal adjustments and ARIMA model-based seasonal adjustments and diagnostic.
References


Daines, M (2005), 'Seasonal adjustment of monetary data: annual review', Bankstats, April, available at www.bankofengland.co.uk/statistics/Documents/ms/articles/art2apr05.pdf.


UNECE (2009), Statistical Metadata (METIS), available at www.unece.org/stats/metis.html.


