Assessing the reliability of monetary statistics

By Chris Wright Tel: 020 7601 4624 Email: chris.wright@bankofengland.co.uk

How reliable are the data in this publication; and what weight should be given to one piece of information relative to another when interpreting economic developments? Answers to these questions are implicit in the impact which the data have on the views of users. But is there more that the statistician can do to inform these judgements? This article reviews some of the factors relevant to the assessment of data quality, and considers how indicators depicting one specific aspect, the reliability of first estimates, might assist users.

What is data quality?

Users of economic statistics place many different, and potentially conflicting, requirements on statisticians. Data must be both accurate and timely; they must be comprehensive and cost effective; and they must be relevant to the particular circumstances of the domestic economy while remaining comparable with data for other countries. The relative importance of these, and other, criteria varies with specific user needs, so that no unique quality ranking can be derived.

Quality assessment thus implies a multi-dimensional approach. As a first step, it requires a set of high level criteria – such as relevance and accuracy - to be defined. Thereafter, tests of compliance with these criteria must be developed. In many cases these tests give rise to qualitative statements rather than a numerical score.

In practice, no unique set of high level criteria has been agreed. For example, parallel studies of data quality undertaken at the International Monetary Fund (IMF)[1] and within the European Commission (Eurostat)[2], led respectively to five and seven high level criteria for data quality assessment (see table 1).

Table 1: Criteria for assessing data quality

Dimension	IMF framework		Eurostat framework
1	Integrity: Firm adherence to the principle of objectivity in the collection, compilation and dissemination of statistics		Relevance
2	Methodological soundness: The conceptual basis for the statistics follows international standards, guidelines and agreed practices		Accuracy
3	Accuracy and reliability: Source data and compilation techniques are sound, and disseminated data sufficiently portray reality		Timeliness
4	Serviceability: Statistics are relevant, timely, consistent, and follow a predictable revisions policy		Accessibility
5	Accessibility: Clear data and metadata are easily available and assistance to users is adequate		Comparability
6			Coherence
7			Completeness

[1] Toward a framework for assessing data quality: IMF Working Paper 25, 2001, by Carol Carson - http://www.imf.org/external/pubs/ft/wp/2001/wp0125.pdf

[2] Quality report prepared for the 14th meeting of the IMF Balance of Payments Working Group: Eurostat, October 2001 - http://www.imf.org/external/pubs/ft/bop/2001/01-42.pdf

Quality criteria

Both the IMF and Eurostat frameworks have been designed for the assessment of entire statistical systems and processes. As a consequence, the high level criteria in Table 1 are intended to address wider questions about the statistical infrastructure, as well as comparisons between families of statistics or individual series. Thus, in developing indicators of "integrity", the IMF are concerned with factors such as the legal and institutional arrangements for data collection; the professionalism of statistical staff; and the independence of statistics from political interference. Similarly, tests of "accessibility" include procedures for the pre-announcement of release dates and standards for the simultaneous release of data to all users, as well as questions concerned with ease of access and availability of methodological notes.

For users faced with interpreting a specific data point, tests of integrity and accessibility may be of secondary importance. Yet there remain a range of criteria on which a judgement is needed. The timeliness and periodicity of data will normally affect the usefulness of the information – what the IMF refer to as "serviceability" – while the consistency of the data with established schemes of classification or accounting frameworks will determine the ability of the user to place the information into context. But, for many users, it is some notion of "accuracy" which is often equated with data quality.

The IMF study develops the linked concepts of accuracy and reliability. Within this framework, accuracy refers to the closeness of the estimated data to the generally unobservable "true" value, while reliability refers to the closeness between initial and final data estimates. In general, the reliability of a compiled statistic increases over time as more information becomes available or as the checking of source data proceeds.

Measuring the accuracy of a series can be problematic. Where data are randomly sampled from a large population, measures of dispersion around the theoretical, but unobservable, true value can be derived. But, if the underlying population is relatively small and highly skewed, or if the source data are volatile from period to period, then standard statistical indicators of accuracy may be more difficult to extract and to interpret.

By contrast, analysing the reliability of data is theoretically straightforward. When data are first released as a preliminary estimate and subsequently revised, eventually to become a final estimate, the size and

direction of revisions are directly observable. Over time, the reliability of the first estimate as a guide to the final data can be expressed both in terms of likely bias or scale of future revisions, based on previous experience. Such information should be helpful to a user seeking to interpret the most timely data estimates.

Revisions and reliability

The distinction drawn here between accuracy and reliability warrants further exploration. As a general statement it is probably reasonable to say that "accuracy" is concerned with the design of the collection system and compilation methodology, while "reliability" concerns its practical implementation.

Revisions to first data estimates can occur for many reasons. In some cases they represent the price users pay for more timely statistics: for example the incorporation of information that was not available to compilers at the earlier release date; or corrections to source data either resulting from the identification of errors or incorporation of more robust raw data estimates. In some other cases, revisions may result from the introduction of methodological changes, which could include: changes which improve the accuracy of the compiled statistics; and/or changes which reduce the cost of their production.

But revisions, or the lack of them, can also result from a policy choice. Some statistics are not revised, or are only revised within the confines of pre-announced guidelines and timetables. A non-revision policy may exist where the statistic is widely used for administrative purposes. The UK Retail Price Index (RPI) is perhaps the best known example of a statistic subject to such a restriction. But a wide range of other macro-economic statistics are open to revision only at predetermined dates and are frozen in other periods, even when changes have occurred within the data used for their compilation. Such restrictions are typically imposed to preserve the coherence of a complex body of statistics, such as the National Accounts, but such an approach is controversial and by no means universally favoured by users.

Reliability indicators must therefore be interpreted with some care. Statistics which score as highly reliable should not automatically be interpreted as "better statistics" than those with a lower score. However, they do provide users with an indication of how much weight to place on a first release of data, that is, the extent to which the first data release can be taken as a reliable indicator of the final – best – estimate of the data to be published at some date in the future.

An application to UK monetary statistics

The remainder of this article reports the results from a preliminary study of data reliability undertaken during the summer of 2001[3]. The objective of the project was to develop possible indicators of reliability of initial estimates for a number of the key monetary series in this publication. The time series examined were monthly changes in non-seasonally adjusted M4 and M4 lending, both in total and by sector. The sectors examined were other financial corporations (OFCs), private non-financial corporations (PNFCs) and the household sector. In addition, retail and wholesale M4 were separately considered, resulting in a total of ten time series. The decision to use non-seasonally adjusted data was to show as clearly as possible the links between revisions and their underlying causes in the source data. By contrast, seasonally adjusted data may be subject to revisions affecting long runs of back data simply as a result of an updating of the mathematical seasonal factors associated with the incorporation of a single new data item. It is for this reason that the regular revisions analyses in the monetary press releases sometimes show extensive changes even though there has been little or no change to the non-seasonally adjusted data.[4]

UK monetary statistics are not subject to any of the restrictions on revisions discussed in the previous section. The data in this publication provide our best estimate of the series concerned at the time of going to press, and new information or improved estimates are incorporated into the data as soon as they become available. One consequence of this policy is that data are never 'final'.

The lack of a final estimate against which to benchmark the first release of data gave rise to the first practical measurement problem. The solution adopted was to trace the evolution of the published data relating to a sample period for each of the selected series. Specifically, for each of our target series, time series were generated showing how the published estimates for each month of 1999 evolved from the first occasion on which data for a particular month became available until a point in the future when estimates for that month's observation appeared to have stabilised. The results from this exercise were analysed and showed that most monthly estimates stabilise about six or seven months after the first estimate, and almost all have done so after nine. Revisions beyond this horizon were typically small, so that estimates nine months after the first release of data were taken to represent 'final' figures.

Adopting this convention allowed the revisions performance of each of the chosen series to be examined in more detail. As a first step, the absolute size of the revision from the first release of data to 'final' data was calculated for each available month in the chosen time series and the results presented as a distribution by size. Data were available from December 1996 to August 2000 for total M4, total M4 lending and retail and wholesale M4. For the sectoral breakdowns of M4 and M4 lending, data were available from November 1997.

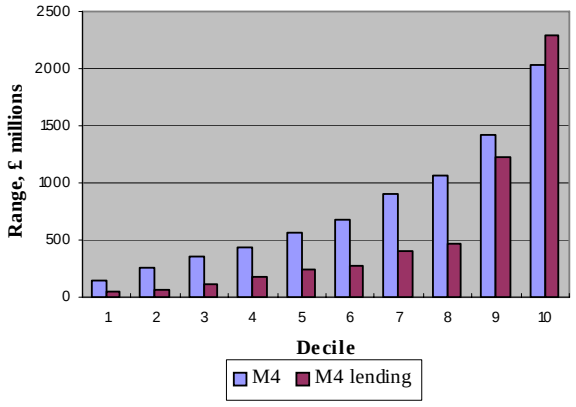
Chart 1 shows the dispersion of the absolute size of revisions for total M4 and M4 lending. The use of deciles to present these data provides a basis for extracting a simple reliability indicator. However, as can be seen, the position chosen for the comparison on the distribution may affect the conclusion drawn. As an example, using

[3] The work was undertaken by Paul Vider, a student at Cambridge University. The remainder of this article, including charts and tables, draws upon his internal report prepared in August 2001.

[4] Storing successive vintages of each month's data has not been a practical proposition in the past, so that the data for this study had to be generated from paper records. This restriction placed practical limitations on the length of the time series that could be constructed and, partly for this reason, the study was restricted to the examination of non-seasonally adjusted data which tend to stabilise earlier than the seasonally adjusted data. In the future it will be possible to generate the revisions history electronically and this should permit an equivalent exercise with seasonally adjusted data to be undertaken.

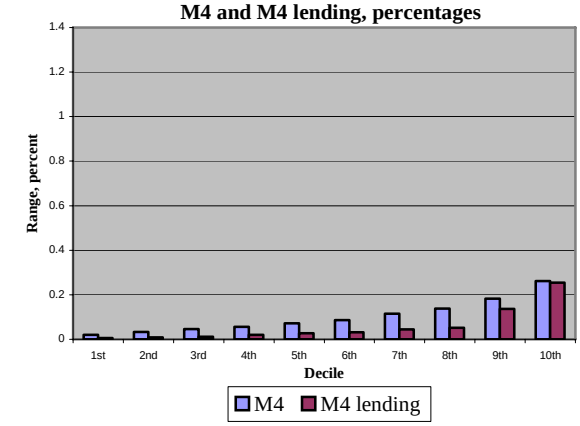
the ninth decile indicates that 90% of first estimates for monthly changes in M4 are within £1420m of the eventual estimate - a slightly larger, although similar, figure to the £1230m for M4 lending. However, if a different point on the distribution is selected, the relative difference in the measured reliability of these two series could be greater – 60% of first estimates of M4 lending are within £280m of the final figure, compared with a much larger range of £670m for total M4. Both comparisons suggest that initial estimates of M4 lending are more reliable than are those for M4 but simple single indicators may only give a partial impression of the relative differences.

Chart 1: Distribution of revisions in decile form to M4 and M4 lending, £ millions



The comparisons in Chart 1 take no account of differences in the scale of the data series under review. Revisions to M4 are typically larger in value than are those for M4 lending, but the stocks of M4 and M4 lending are different in size, so that comparisons using revisions expressed as a proportion of their outstanding stock provide a clearer picture of relative reliability. Chart 2 represents the data in this format. As can be seen, M4 lending is confirmed as being relatively more reliable than M4 – first estimates proportionately closer to final values – but again with the rider that the comparison is sensitive to the selected position on the distribution of revisions.

Chart 2: Distribution of revisions in decile form to M4 and M4 lending, percentages



Within the totals there was considerable variation in the results for different sectors (see Charts 3 and 4). For example, in Chart 3, OFCs' M4 holdings were typically subject to the largest revisions, with those at the ninth decile equivalent to as much as 0.76% of the outstanding stock, compared to 0.49% for PNFCs and just 0.18% for households. The greater reliability of first estimates for the household sector's M4 holdings, on this measure, is in part a reflection of the scale of the sector's holdings. The average size of revisions, expressed as absolute values, was indeed highest for OFCs, but the value of revisions to the household sector's holdings was typically larger than those for PNFCs while remaining much smaller as a percentage of the outstanding stock (see Table 2). The greater reliability of first estimates of data for households over other sectors is repeated for M4 lending (see Chart 4). Moreover, first estimates for lending to both OFCs and households is more reliable than their M4 equivalents.

Chart 3: Distribution of revisions in decile form of sectoral breakdown of M4, percentages

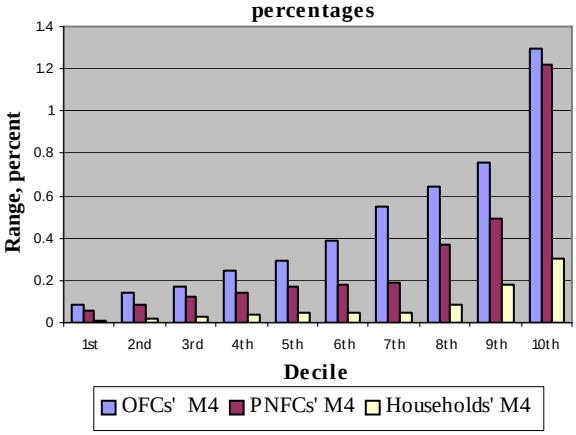
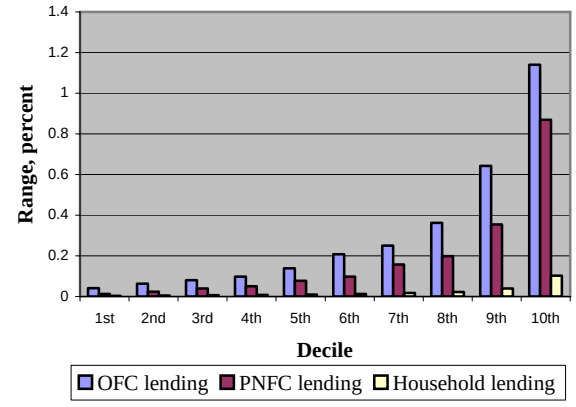


Chart 4: Distribution of revisions in decile form of sectoral breakdown of M4 lending, percentages



Revisions to retail and wholesale M4 were of similar magnitude in value terms - 90% of revisions within £1050m and £1070m respectively. However, when allowance is made for the larger stock of retail holdings, the first estimate for this series is shown to be the more reliable (see Chart 5).

The revisions data, expressed in value terms, were also used to examine the possible presence of bias within first estimates of the monetary statistics. An initial graphical examination of the data suggested that revisions to first estimates may be broadly normally distributed with an average close to zero, indicating little or no significant bias. This impression is confirmed more formally in

Table 3. This shows the mean value of revisions for each data series together with standard deviations. While the means can differ from zero by more than £100m, the calculated standard deviations are substantially greater, reflecting the very wide range of revisions, positive and negative, revealed within the data. The bottom row of the table shows the T ratios for each series – mean over standard deviation. For bias to be present, i.e. for the mean value of all revisions to be different from zero, the sample mean would generally be expected to be at least two standard deviations from zero. In practice, none of the recorded T values is higher than 0.34 suggesting that the risk of systematic bias in the ten series investigated is very low.

Table 2: 9th decile of revisions as an absolute value and as a percentage of the total stock

	M4	M4 lending	Retail M4	Wholesale M4	OFC M4
Range, £ million	1418	1226	1047	1069	1413
Stock, £ million	778,548	896,684	510,472	268,066	186,766
As % of stock	**0.182**	**0.137**	**0.205**	**0.399**	**0.756**

	PNFC M4	Household M4	OFC lending	PNFC lending	Household lending
Range, £ million	590	901	1272	690	206
Stock, £ million	120,328	489,711	198,009	194,435	530,095
As % of stock	**0.490**	**0.184**	**0.642**	**0.355**	**0.039**

The study also examined the speed at which successive data estimates converged on their final values. Table 4 shows the percentage of the total nine month revision achieved in the first three months. In the first row of the table, a figure of 100% would indicate that the data had achieved their final value within three months. Figures below 100% provide an indication of the rate of adjustment while figures above 100% show that early revisions caused the data to overshoot their final value.

Table 3: Mean and standard deviations of revisions for all data series, £ millions

	M4	M4 lending	Retail deposits	Wholesale Deposits	OFC holdings
Average, £ millions	115	149	-1	114	308
Standard deviation £ millions	841	614	532	787	898
Avg/SD	0.1364	0.2428	-0.0019	0.1454	0.3426

	PNFC holdings	Household holdings	OFC lending	PNFC lending	Household lending
Average, £ millions	-123	-33	214	-3	-14
Standard deviation £ millions	420	453	701	520	139
Avg/SD	-0.2919	-0.0722	0.3051	-0.0058	-0.0990

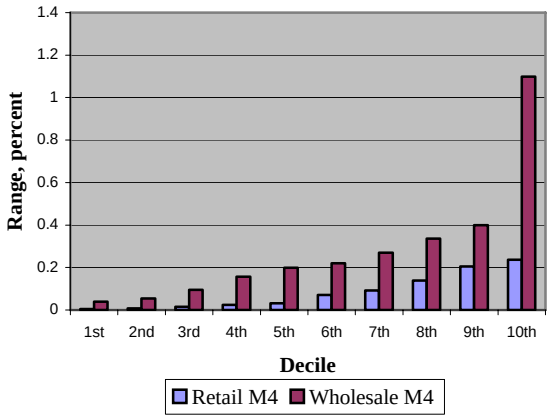
Table 4: Average percentage of total nine-month revision completed after three months

	M4	M4 lending	Retail M4	Wholesale M4	OFC M4
Average, %	114.0	86.1	136.4	83.3	169.2
Standard deviation, %	148.3	232.3	365.6	206.5	657.9

	PNFC M4	Household M4	OFC lending	PNFC lending	Household lending
Average, %	2.5	56.1	72.4	73.3	77.2
Standard deviation, %	252.9	107.6	79.0	150.1	82.3

The results confirmed that data estimates three months after their first release were more reliable than the initial estimate in the majority of cases. On average 86% of the total revision to M4 lending had been achieved just three months after the first estimate, and all sub-sectors had completed over 70% of their total revision. By contrast, the results for M4 holdings were more erratic. Although total M4 (like total lending) was typically very close to its final value, it showed a tendency to overshoot. The table shows that this was explained by sizeable over-adjustment of estimates for holdings by OFCs – more than offsetting the very limited adjustments for the other two sub-sectors. However, the associated standard deviations for these average adjustments are large. Probably the best one can say is that by the third month the reliability of the estimates had improved and that M4 lending converges on its final estimate more quickly and more consistently than M4 holdings.

Chart 5: Distribution of revisions in decile form of retail and wholesale of M4, percentages



Conclusions

More work will be needed to develop a range of reliability and wider quality indicators for monetary statistics. But the preliminary results reported here are considered to be an important first step. The study has shown that first data estimates for the selected series are unbiased and that the M4 lending aggregate and components are relatively more reliable than the equivalent estimates for M4 holdings. This came as no surprise because the compilation of M4 relies on the incorporation of data not

reported directly to the Bank, and not available to the same timetable. This is particularly true for data on the sectoral holdings of marketable instruments issued by banks and building societies, because the issuers are not able to report the holders of such instruments. This aspect of the compilation of M4 is examined in more detail in the article entitled *Compilation methods of the components of broad money and its balance sheet counterparts*, also in this publication; see especially the section on wholesale deposits. Within the sectoral data, lending to, and holdings by, households are generally estimated more reliably at the time of the first release of data than are other sectors.

Future work will begin to examine the properties of seasonally adjusted data but will also start to document and report a wider range of information to assist users. Feedback from users on the preferred form and content of such guidance will assist the planning of this work.