

I. INTRODUCTION

Are the figures accurate? Are these data reliable? Questions like these are important to the producers and users of any set of statistics. Getting answers is never easy, but is particularly difficult in the case of datasets such as national accounts and balance of payments. Both of these datasets are based on multiple, complex source data and typically undergo several routine revisions as more and better source data are incorporated into the final estimates. As a result, neither national accounts nor balance of payments statistics can be subjected directly to the usual statistical measures of sampling biases, variances, and other measurement error properties. In this context, this note explores approaches that have been used to shed light on the accuracy and reliability of these datasets.

This note takes as its frame of reference the dimension of the IMF's Data Quality Assessment Framework (DQAF) that deals with accuracy and reliability. It focuses on a subset of the elements and indicators in that dimension—namely those concerned with assessments of statistical outputs (as distinguished from assessments of source data or intermediate data). The purposes of the note are at least three-fold: (1) to increase awareness of some of the approaches to assessing the quality of statistical output that are used in national accounts and balance of payments, (2) to stimulate thinking about additional approaches that might be used to assess accuracy and reliability in such datasets, and (3) to stimulate thinking about how these approaches can be generalized to other datasets. Thus, the note is addressed not only to those interested in national accounts and balance of payments, but also to those interested in a range of other datasets.

The note does not attempt to break new ground. Rather it pulls together examples of the several approaches currently used to assess statistical outputs in national accounts and balance of payments, and it sets them out in a way that facilitates consideration of their use in a wider range of countries and datasets. The examples are drawn from material put in the public domain over roughly the last decade, but the note makes no pretense of being exhaustive or representative.

The structure of the note is as follows. Section II sketches the Accuracy and Reliability dimension of the DQAF and the underlying definitions of accuracy and reliability. Section III describes several approaches to assessing accuracy and reliability used in national accounts and balance of payments, giving some examples from statistical agency documents. Section IV invites responses and comments. Section V is a postscript about the application of the approaches presented in the note as pertaining to statistical outputs to earlier stages of the statistical process. A bibliography identifies primary sources of general interest on the topic, while footnotes focus on sources of the specific examples.

II. THE ACCURACY AND RELIABILITY DIMENSION OF THE DQAF

It is increasingly commonplace to say that statisticians used to equate the quality of statistics with accuracy, reliability, or both—sometimes these words were used as meaning the same thing and sometimes not. In the recent upswing in attention given to data quality, in which

the work toward the DQAF played a central role, data quality is taken to be multi-dimensional, and accuracy and reliability take their place within the longer list of the dimensions of quality.² As the DQAF has taken shape, accuracy and reliability were put together in a single dimension, which is built around the following definitions.

- **Accuracy** refers to the closeness between the estimated value and the (unknown) true value that the statistics were intended to measure. Assessing the accuracy of an estimate involves evaluating the error associated with an estimate. In practical terms, there is no single aggregate or overall measure of accuracy; accuracy is evaluated in terms of the potential sources of error, and these potential sources of error differ across datasets.
- **Reliability** refers to the closeness of the initial estimated value(s) to the subsequent estimated values.³ Assessing reliability involves comparing estimates over time. In other words, assessing reliability refers to revisions. This feature is identified separately for two reasons. First, it is usually the initial estimate that captures attention, whence the importance of its accuracy. Second, the separation helps bring out the fact that data that are not revised are not necessarily the most accurate.

The Accuracy and Reliability dimension from the DQAF generic framework, which is applicable to both national accounts and balance of payments (as well as to other datasets), is reproduced in Box 1.⁴ The so-called “pointers to quality”—that is, the elements and indicators (in the second and third columns of the box)—are largely familiar to statisticians. These pointers could have been arranged in several ways. For example, all of them at a stage in the statistical process could have been grouped together. With such a scheme, pointers related to data collection, data processing, through to the final output data, and dissemination would have been grouped separately. As now arranged in the dimension, they are grouped pragmatically. First, elements 3.1 and 3.2, Source Data and Estimating Techniques, respectively, represent the basic determinants of accuracy and reliability. Second, elements 3.3, 3.4, and 3.5 all relate to processes, and they focus on source data, intermediate results,

²For information about the DQAF, see Fourth Review of the Fund’s Data Standards Initiative: Supplement on the Data Quality Assessment Framework, on the IMF website at <http://www.imf.org/external/np/sta/dsbb/2001/supp.htm>.

³The term to be used for this characteristic has been elusive. However, because the term “reliability” is firmly embedded in the literature about revisions from the United States, Australia (along with “revisability”), and the United Kingdom, it was adopted for the DQAF for consistency with that considerable body of literature.

⁴The DQAF comprises a generic framework, which is applicable to all datasets, and a set of specific frameworks that provide more detail and concreteness tailored to the dataset in question.

and statistical outputs. Element 3.3 deals with the assessment and validation of source data, with attention to survey methodology in particular. Element 3.4 deals with assessment and validation of intermediate data and statistical outputs. Element 3.5 also deals with statistical output, specifically on revisions in output data.

**Box 1. The Accuracy and Reliability Dimension of the Data Quality
Assessment Framework—Generic Framework
(As of July 2001)***

Quality Dimensions	Elements	Indicators
<p>3. Accuracy and reliability</p> <p><i>Source data and statistical techniques are sound and statistical outputs sufficiently portray reality.</i></p>	<p>3.1 Source data – <i>Source data available provide an adequate basis to compile statistics.</i></p> <p>3.2 Statistical techniques – <i>Statistical techniques employed conform to sound statistical procedures.</i></p> <p>3.3 Assessment and validation of source data – <i>Source data are regularly assessed and validated.</i></p> <p>3.4 Assessment and validation of intermediate data and statistical outputs – <i>Intermediate results and statistical outputs are regularly assessed and validated.</i></p> <p>3.5 Revision studies – <i>Revisions, as a gauge of reliability, are tracked and mined for the information they may provide.</i></p>	<p>3.1.1 Source data are collected from comprehensive data collection programs that take into account country-specific conditions.</p> <p>3.1.2 Source data reasonably approximate the definitions, scope, classifications, valuation, and time of recording required.</p> <p>3.1.3 Source data are timely.</p> <p>3.2.1 Data compilation employs sound statistical techniques.</p> <p>3.2.2 Other statistical procedures (e.g., data adjustments and transformations, and statistical analysis) employ sound statistical techniques.</p> <p>3.3.1 Source data—including censuses, sample surveys and administrative records—are routinely assessed, e.g., for coverage, sample error, response error, and non-sampling error; the results of the assessments are monitored and made available to guide planning.</p> <p>3.4.1 Main intermediate data are validated against other information where applicable.</p> <p>3.4.2 Statistical discrepancies in intermediate data are assessed and investigated.</p> <p>3.4.3 Statistical discrepancies and other potential indicators of problems in statistical outputs are investigated.</p> <p>3.5.1 Studies and analyses of revisions are carried out routinely and used to inform statistical processes.</p>

*Other quality dimensions are 1. Integrity, 2. Methodological Soundness, 4. Serviceability, and 5. Accessibility; O. Prerequisites of Quality cover overarching characteristics. See <http://www.imf.org/external/np/sta/dsbb/2001/supp.htm>.

It should be noted that the focus of these elements is not so much on the results the assessments yield as on the practice of conducting such assessment and validation exercises. Thus, the pointers that deal with assessment and validation in this dimension of the DQAF query the existence of such processes and their use to inform the statistical processes and guide planning of the statistical agency. They are to be distinguished from those about informing of the results of the assessment processes, as covered in an element in the Accessibility dimension of the DQAF, which deals with making appropriate information available to the public.

III. APPROACHES TO ASSESSING STATISTICAL OUTPUTS IN NATIONAL ACCOUNTS AND BALANCE OF PAYMENTS

National accounts and balance of payments are both the subject of this note because the two systems are alike in that they are based on multiple, complex source data and typically undergo several routine revisions as more and better source data are incorporated into the statistical outputs.⁵ These features have important implications for the assessment of data quality. One implication is that, while the standard measures are applicable at the level of specific data sources and thus have a role to play in the overall assessment of data quality, these standard measures are not directly applicable to statistical outputs of national accounts and balance of payments. This situation is well recognized:

In principle, the quality characteristics of national accounts aggregates should be rigorously ascertained from measured sampling biases and variances, and other measurement error properties of the input data. In practice, this is not possible, certainly not in any comprehensive manner, given the complexity of the estimation methods involved, the variety of input and the lack of reliable measures of error for many of these. Simon Kuznets, the distinguished American statistician, expounded the argument some fifty years ago: "To analyze the reliability of data and procedures used to derive national income total and their components is essentially an insoluble task." [Statistics Canada, 1990, p. 104.]

Compilation of balance of payments...is a complex task. Given the variety of sources and methods used, there is no single comprehensive measure of the quality of these estimates. Nevertheless, each of the quality indicators described below provides a partial insight into the quality of the statistics. To get an overall picture, all measures need to be viewed together while taking account of their limitations. At best such an assessment can only be subjective. [ABS, 1998, paragraph 15.15.]

⁵Much of what is said about assessing accuracy and reliability of balance of payments may also apply to international investment position statistics.

The following subsections discuss four approaches that can be used in assessing accuracy and reliability of statistical output: examination of statistical discrepancies, comparison with other data, analysis of revisions, and judgmental evaluation.⁶ The evaluations of the approaches that are mentioned, unless otherwise noted, are drawn from the original sources. Some case studies of the applications of these approaches are also included.

A. Statistical Discrepancies

Statistical discrepancies may be defined most generally as the difference between two totals that **should** be equal. The size, sign, and variability of these discrepancies may shed some light on accuracy, and indeed may suggest that something is amiss in the output data.

In **national accounts**, the statistical discrepancy most widely viewed as serving this role is the difference between the sums of the components that add up to GDP derived from the income, product, and by-industry methods of measurement. In the U.S. national income and product accounts (NIPAs) prepared by the Bureau of Economic Analysis (BEA), the statistical discrepancy, in current dollars, is shown on the income side of the national income and product account where it is the difference between the sum of final expenditures and inventory change and the sum of the "income" components. BEA publishes the "statistical discrepancy" in NIPA tables in the *Survey of Current Business* (tables 1.9 and 5.1). This discrepancy reflects error on both sides of the account, but for two reasons it is not a good measure of error. First, some errors on the two sides are not independent. Second, in preparing estimates, BEA makes adjustments that limit the size of the discrepancy, especially in the quarterly estimates.

The statistical discrepancy has been drawn into recent discussion of the accuracy of the U.S. GDP estimates. Over the 10 years before the 1995-96 benchmark revision, the statistical discrepancy averaged 0.3 percent of GDP. In 1996-98, the average was still about that size, but in 1999 and 2000 the sign changed (it became negative) and it averaged 0.8 percent of GDP. The Council of Economic Advisers, in *The Economic Report of the President* (February 1997), expressed concern over the size of the discrepancy and, drawing on analysis of some relationships, concluded that the product-side measure was understating growth.⁷

⁶Another approach, which may be seen as a variant of time series analysis, has not been included. For example, one evaluation of "noise-to-signal" ratios calculated for GDP aggregates using an X11-ARIMA seasonal adjustment model was identified, but it does not seem to have been followed up [see Statistics Canada, 1990, p. 101]. Another example of this approach is the reconciliation model for intra European trade statistics, which is still at an experimental stage [see Tyrman, 2001].

⁷(Washington, D.C.: U.S. Government Printing Office), pp. 72-74.

BEA responded with its reasons why the product-side estimates were stronger than the income-side estimates and noted its commitment to continue to work to reduce the size of the discrepancy.⁸

For the **balance of payments**, the focus is on the net errors and omissions item. The use of the double-entry accounting system of recording means that, in principle, the net sum of all credit and debit entries should equal zero. In practice, such equality rarely exists, and any differences are recorded in the net errors and omissions item. The Australian Bureau of Statistics (ABS) makes the following comments about the use of this item:

Persistently large figures in one direction (negative or positive) may be taken as an indication of serious and systematic errors. However, a small figure does not necessarily mean that only small errors and omissions have occurred, since large positive and negative errors may be offsetting. Offsetting errors may be either related or unrelated, resulting from a measurement problem affecting either both sides or only one side of a transaction. If positive and negative net errors and omissions tend to offset each other in successive periods, errors may be due to time differences in data reported by different sources to estimate the credit and debit sides of a transaction. [ABS, 1998, paragraph 15.18.]

The following example may be cited to show how the analysis of the statistical discrepancy has been used. About a decade ago, the large and persistent statistical discrepancy of the same sign in the U.S. balance of payments raised concern about whether the estimates were capturing reality. In 1990, the discrepancy between the current account and the capital account (now referred to as the financial account) was reported to be unprecedentedly large (US\$73 billion). This size was particularly troubling at that time because, after a decade of large recorded net capital inflows, lower rates of return and increased uncertainty about the U.S. economy appeared to have, combined with increased demand for credit in the rest of the world, reduced the supply of capital to the United States. The resulting large drop in recorded net capital inflows was not matched, however, by a similar drop in the current account deficit. If the current account was correct, the United States must still have been borrowing large sums from abroad to finance its deficit in goods, services, income, and unilateral transfers. The large statistical discrepancy made it difficult to determine whether the supply of foreign capital had indeed been reduced. [BEA, 1995.] Subsequently, substantial efforts were made to strengthen the source data for the capital account.

B. Comparisons with Like Estimates

Comparison of estimates that purport to measure the same or related phenomena but are drawn from different sets of statistics—even after reconciliation of concepts, time of

⁸"The Statistical Discrepancy" box in the *Survey of Current Business*, August 1997, p. 19.

recording, valuations, etc.—may shed light on quality or suggest that something is amiss. This process, sometimes called data confrontation, is very familiar to statisticians because it is carried out not only for output data (which is the subject of this note), but also for source data and intermediate data. As well, it may be familiar to data users, who may attempt it when faced with puzzling statistics. For output data, such comparisons take a number of forms. Although the distinctions among the forms tend to blur, they may be grouped as (1) comparisons with different sources, (2) comparisons of corresponding components of different macroeconomic datasets, and (3) comparisons of partner country data.

Comparisons with different sources

For its **national accounts**, BEA publishes a number of "relation" tables that show the coverage, valuation, timing, and other sources of difference between the NIPA estimate and other estimates.⁹ A number of the tables refer to types of income, and they relate BEA's published estimates to another agency's data. For example:

- Relation of Consumption of Fixed Capital in the National Income and Product Accounts to Depreciation and Amortization as Published by the Internal Revenue Service.
- Relation of Corporate Profits, Taxes, and Dividends in the National Income and Product Accounts to Corresponding Measures Published by the Internal Revenue Service.
- Relation of Net Farm Income in the National Income and Product Accounts to Net Farm Income Published by the U.S. Department of Agriculture.
- Relation of Wages and Salaries in the National Income and Product Accounts to Wages and Salaries as Published by the Bureau of Labor Statistics.

As well, from time to time BEA publishes articles that compare NIPA estimates with other measures. For example, a recent article compares NIPA profits with Standard and Poor's operating earnings.^{10, 11}

⁹The most recent publication of the annual tables was in the August 2001 *Survey of Current Business*, in the "National Income and Product Tables."

¹⁰Kenneth A. Petrick, "Comparing NIPA Profits with S&P 500 Profits," *Survey of Current Business*, April 2001, pp. 16-20.

¹¹Another approach that may be seen as a combination of source data and assessment technique is the use of an events tracking system to monitor international transactions by scanning business media. The tracked events are both company-specific and of a general
(continued...)

Comparisons of corresponding components of different macroeconomic datasets

In the context of this note, it would be logical to think of comparisons of **national accounts and balance of payments**. BEA, for example, publishes each month a table that shows a reconciliation of the NIPA net exports of goods and services and net receipts of income to the balance on goods, services, and income in the International Transactions Accounts (balance of payments). A more detailed comparison is provided in an annual table that shows the relation between foreign transactions in the NIPAs to corresponding items in the International Transactions Accounts. Separately for exports and imports of goods, for exports and imports of services, for receipts and payments of income, and for net unilateral transfers as well as the balances, the table identifies geographical differences in coverage, conceptual differences, and statistical differences (revisions incorporated in one dataset but not the other).¹²

Comparisons of partner country data

For **balance of payments**, partner country comparisons are conducted in both bilateral and multilateral settings. The comparisons are based on the principle that an outflow (inflow) from one country to another country should be recorded as an inflow (outflow) for that other country.

The United States and Canada have long conducted an annual exercise to compare their current account estimates, and for the last decade the co-authored results have been published.¹³ Bilateral reconciliations are also conducted by a number of other countries—for example, between Australia and New Zealand on goods and services trade and on capital flows, and between Australia and several partners—the United States, Japan, and the European Union—on goods trade. [ABS, 2001.]

Multilateral reconciliations may be represented by those conducted under the auspices of the Asia Pacific Economic Cooperation (APEC) initiative. A database was created to analyze available data in the fields of goods and services trade and international direct investment flows for the 17 member countries participating. Efforts were made to identify areas where major differences existed in partner country data, with the aim of reducing these differences

background nature. The information is used by Statistics Canada in editing and updating of survey coverage for the balance of payments. [Statistics Canada, no date, p. 2.]

¹²The most recent publication of this table was in the August 2001 *Survey of Current Business*, Table 4.5 B in the “National Income and Product Tables.”

¹³The latest available reconciliation was published in the November 2001 *Survey of Current Business* and in *Canada's Balance of International Payments, Third Quarter 2001*.

over time. [APEC, 1998.] Also, multilateral reconciliations are now undertaken in the European Union and Euro area contexts.

Partner-country data comparisons were a key part of the study undertaken in the early 1990s by the IMF's Working Party on the Measurement of International Capital Flows, and its report has continued to shape international efforts to improve balance of payments data. The Working Party undertook data confrontation using the financial account data of various countries, and reported on various specific examples (e.g., for reinvested earnings on direct investment). It concluded that "national data providing geographic details of (financial) flows vis-à-vis partner countries...are particularly useful in detecting and quantifying gaps and discrepancies." [IMF, 1992, p. 33.] Flowing from that conclusion, the Working Party recommended that countries collect stock and flow data on a country-by-country basis and exchange these data. Another related recommendation was that countries engaged in significant amounts of investing should conduct a coordinated survey of portfolio investment positions, broken down by partner country, to enhance coverage, to ensure uniformity of data reporting practices, and to serve as a benchmark for addressing gaps in the reporting of portfolio investment flows. One outcome was the Coordinated Portfolio Investment Survey. The survey facilitated bilateral comparisons of national data on portfolio investment assets and liabilities; some examples are cited in an analysis of the results of the 1997 survey.¹⁴

C. Revisions

The analysis of revisions is a means of assessing the quality of the first (or other relatively early) estimate in relation to later, sometimes, final estimates. One method is to prepare and examine measures of revision. These studies are typically done by statistical agencies and draw on databases that record the various vintages of estimates. As well, there are several other methods, including use of statistical models.

Measures of revision

The studies of measures of revision are basically similar, but vary somewhat in at least three ways: (1) the vintages of estimates that are compared, (2) the component detail in which they are conducted, and (3) the measures of revision they yield.

BEA has done studies of revisions of **national accounts** estimates, including GDP and selected other aggregates, for several decades. The current series of studies is undertaken in response to the requirement that it prepare, under Statistical Policy Directive No. 3, an evaluation every three years of the "accuracy and reliability" of the GDP estimates. In fulfilling the requirement, BEA writes as follows:

¹⁴International Monetary Fund, 2000, *Analysis of 1997 Coordinated Portfolio Investment Survey Results and Plans for the 2001 Survey* (Washington: IMF).

...[BEA] evaluates GDP performance using measures of revisions. It does not directly address the “accuracy” of GDP, because such an evaluation would require data on the total measurement error, which cannot be observed. This total error arises from errors in the source data and in the estimating procedures that use the source data. Assuming that later estimates are more accurate than earlier ones, the revisions reflect improvements in accuracy relative to earlier estimates, although the later estimates may contain unknown errors. [Grimm and Parker, 1998, p. 13.]

BEA has prepared estimates of dispersion (mean of absolute values of the revisions), relative dispersion (dispersion as a percentage of the average of the absolute values of the latest estimates), and bias (mean of the value of the revisions).¹⁵ In the latest published article [Grimm and Parker, 1998], it emphasized dispersion because (1) the bias estimates are generally small and (2) small changes in the time period examined often result in substantial changes in the measures of bias.¹⁶ Such measures, BEA notes, must be used with several considerations in mind. Focusing on aggregates, such as GDP, overemphasizes the role of the accounts as providing summary aggregates and downplays their role in displaying the interactions of sectors. Emphasizing revisions puts a penalty on making improvements because improvements are causes of revisions.

The U.S. studies of revision show that the current quarterly estimates (that is, the estimate released near the end of the first, second, and third month after the end of the quarter) of GDP indicate the following:

¹⁵As typically calculated using percent change from quarter to quarter, at annual rates

dispersion is: $\Sigma |P - L| / n$,

where P is the percentage change in the current estimate, L is the percentage change in the latest estimate, n is the number of quarterly changes,

relative dispersion is: $\frac{\Sigma |P - L| / n}{\Sigma |L| / n}$, and

bias is: $\Sigma(P - L) / n$.

¹⁶A new study is expected (as of December 2001) in a forthcoming issue of the *Survey of Current Business*.

- whether the economy is expanding or contracting,
- whether growth is accelerating or decelerating,
- whether the growth rate is high or low relative to trend.

However, the estimates' ability to do this is least when growth is hovering near zero and—although the evidence is less clear—at turning points in the economy. The quarterly estimates identified 4 of the 5 peaks, and 3 of the 5 troughs, between 1969 and 2000. In all of the misses, the miss was by one quarter. [Grimm and Parker, 1998, pp.12-13.]

For **national accounts**, the U.K. Office of National Statistics has carried out a series of revisions studies beginning in the early 1990s.¹⁷ The most recent study covered constant price GDP growth (and its components) since 1970. It dealt with annual estimates between successive editions of the *United Kingdom National Accounts* (“the Blue Book”), namely the edition in which the estimate appears the first time and that in which it appears the second time. (It does not consider revisions to the first publication of the initial estimate of annual GDP growth and the initial publication in the Blue Book.) The study focused on bias; bias “provides information about the reliability of a series, but not about the accuracy of a series.” [Symons, 2001, p. 41.] It also considered the dispersion of the revision and the mean square error, the latter because it captures the notions of bias and dispersion of revisions in one measure. (The mean square error of a series is the sum of the square of the bias and the variance of the series.)

Among the results of the analysis were the following points:

- A mean upward revision in the initial estimate of GDP growth of 0.2 percentage points over the past 29 years compared with average annual growth of 2 percent over this period.
- The revision to the initial estimate is statistically biased.
- The mean revision and dispersion of revisions to the initial estimates of GNP growth is lower than many of its components.
- The mean revision and dispersion of the revisions to the initial estimates of GDP growth have fallen in each decade over the past 30 years. [Symons, 2001, pp. 41-42.]

For **balance of payments**, the ABS also focuses on measures of bias and dispersion, which it characterizes as follows: bias is a measure of the extent to which the initial estimate is generally higher or lower than the latest estimate, and indicates the direction of the revisions.

¹⁷The studies covered, in addition to national accounts, other “headline” series such as the index of production and the balance of trade in goods; [see Barklem, 2000].

Dispersion is a measure of the spread of latest estimates about the initial estimate, and indicates the magnitude of the revisions. The ABS notes several general points that need to be kept in mind when considering revisions:

- Revision studies reflect past experience and may not be a good indicator of behavior in the immediate future.
- The findings for aggregates for net series should be treated with caution because they reflect varying impacts of revisions of their components.
- Substantial change in the volume of transactions for an aggregate occurring over a relative short time will make the effect on revision difficult to predict.
- The latest estimate is assumed to be a better approximation of the notional true value than earlier estimates.
- Significant changes in a concept or methodology need to be considered carefully (e.g., a better approximation of a concept may be an improvement, but it will impair the record of revisions unless the impact is isolated. [ABS, 1998, paragraph 15.29.]

Such studies of revision in national accounts have been done in several countries in a similar enough way that an occasional attempt has been made to compare the measures. For example, in addition to the measures for Australia, Canada, and the United States, a study in 1990 included measures for Germany and Japan.¹⁸

Other revision studies

One additional approach to revision analysis deserves mention not only because it is attractively “low-tech” but also because it can be used to shed light on the sources of the revisions. As applied in a study of the global capital account (now referred to as the financial account) discrepancy, the approach consists of laying out a series of four matrix tables, with reference dates (e.g., 1998, 1999, 2000) in the column headings and vintages of the estimates in the row headings. The tables were for (1) the values of a variable, (2) revisions in the variable (first differences derived from the first table), (3) the revisions due to a source of revision (such as a methodological change), and (4) the other (“normal”) revisions. The approach was used to answer the question whether revisions tend to reduce imbalances in the global accounts. [IMF, 1992, Appendix VIII.]

Another approach to studies of revision is to focus on the impact of revisions on users. One such study dealt with the revisions of the national accounts estimates to ask to what extent the early estimates might have misled policy makers. In a paper prepared in 1986, the authors

¹⁸[de Leeuw, 1990, pp.191-203.]

reported that policymakers had stated that preliminary estimates had been misleading in six instances; they agreed in four cases but in the other two did not find a strong case for believing that the estimates had been misleading. [Carson and Jaszi, 1998.]

In the United States, a number of studies of revisions of the national accounts have been prepared outside BEA. Several use statistical techniques—such as an errors-in-variables model, a rational forecast model, a generalized method of moments, and tests for co-integration and stability—to analyze revisions. [These studies are summarized in U.S. BEA, 1995.] BEA concluded that their principal implication is that "some improvements could be made in the early estimates. Despite several limitations, these studies have provided BEA with tools to further evaluate its revisions."¹⁹

D. Judgmental Valuation

The approaches to the assessment of the accuracy and reliability of output data just mentioned all find a place in the Data Quality Assessment Framework (DQAF), in elements 3.4 and 3.5. Some national agencies also identify another approach, referred to as judgmental (or subjective) evaluation. It is not included in the DQAF's Accuracy and Reliability dimension because, in effect, it can be regarded as a synthesis of what experts might conclude on the basis of many or all the dimension's elements. It is included in this note for completeness and because some of the points made are worth keeping in mind when assessing quality.

The assumption underlying these evaluations is that it is possible, from knowledge of the data, to form very rough and mainly subjective judgments of the ranges of reasonable doubt attaching to the estimates. The evaluations have been applied to components, and they use three or four "grades." They vary as to whether they evaluate the initial or final estimate. They also vary somewhat in the statements of the grounds for the evaluations.

The Canadian evaluation of **national accounts** lists the factors that underlie a three-grade "subjective quality assessment" of the final estimates of current price components of GDP:

1. Most reliable: (a) estimates are based on highly reliable sources and (b) the concepts and definitions underlying the input data closely correspond to those required or adjustments are straightforward.
2. Reliable: sources are administrative records or surveys that are not highly reliable or require difficult, error-prone adjustments.
3. Acceptable: direct, reliable observation is not possible and therefore the estimates depend on judgment to a large degree or are based on related indicators.

¹⁹[U.S. BEA, 1995, p. 55.]

The rating of constant price components depends on the same features; a reduction in quality, if any, is attributable to corresponding price indexes. [Statistics Canada, 1990.]

The ABS prepares ratings for the current price income and expenditure components of GDP, for the chain volume measures of the expenditure components of GDP, and for the industry value added chain volume measures. The ratings pertain to the initial quarterly estimates of movement of key components; initial quarterly estimates of movement have been chosen as they are generally the most anticipated of the national accounts estimates. The ratings are A (good), B (fair), C (poor), and D (very poor). The presentation opens with the following statement:

While it is generally not possible to provide exact information on the accuracy of national accounts estimates, intuitive assessment of the accuracy of the estimates can be made, based on knowledge of data sources used. [ABS, 2000, paragraphs 29.18-29.22.]

For **balance of payments**, Statistics Canada uses the same three levels as for national accounts—most reliable, reliable, and acceptable. The indicator of accuracy is applied to each specific account of the balance of payments. It is seen as representing “the professional judgment of statisticians as to the degree of error and bias taking into account the available sources of information and the methodology used.” [Statistics Canada, 2000, p. 2.]

The ABS provides a rating of the principal balance of payments components: A (less than 5 percent margin of error); B (5 percent to less than 10 percent); C (10 percent to less than 15 percent); D (15 percent or greater). The ratings apply to the initial quarterly and annual estimates. The ratings are assessments of the quality of the estimates in terms of (1) the possible discrepancy between the estimates value and the true value and (2) the upper bounds in which revisions may occur from time to time. The assessments are based on the following factors:

- Analyses of the statistical processes within the agency,
- Observation of the types of error occurring,
- Examination of the residual and of consistency in the behavior of series,
- Comparisons with partner country data, and
- Revisions history of the series.

The ratings are shown in a table along with a recent value to give an idea of the relative importance of the item. [ABS, 1998, paragraph 15.30.]

IV. INVITATION TO RESPOND AND COMMENT

By design, the national accounts and balance of payments are frameworks that integrate in a systematic fashion a wide range of statistics obtained from myriad of data sources made available at various time periods. As such, the assessment techniques, such as those based on survey methodology, that are widely used on stand-alone statistics are not directly applicable to statistical outputs of national accounts and balance of payments. Four approaches were reviewed in this note: statistical discrepancies that exploit the integrated nature of the dataset; comparison with other data, such as like data from other sources, confrontation with corresponding components of different datasets, and assessment against partner country data; analysis of revisions, and judgmental evaluation that draws on the expertise and in-depth knowledge of statisticians.

The questions that follow, first, invite the reader to contribute additional examples of the approaches discussed in the paper and their application. The goal would be to assemble a reference document to assist those interested in assessments of accuracy and reliability. Also, several questions follow up on the second and third purposes of this paper: to stimulate thinking about additional approaches that might be used in national accounts and balance of payments and about how these approaches can be generalized to other datasets.

1. The note lists four main approaches to assessing accuracy and reliability of national accounts and balance of payments. Are there approaches that have not been identified? Should time series analysis, mentioned in footnote 6, be explored more fully?
2. The examples are drawn from a limited set of country experiences. Are there additional examples, particularly those that are more up to date and/or are available on the Internet, that could be usefully cited?
3. The paper mentioned several examples where the approaches proved useful in signaling that a problem exists and sometimes pointing toward a remedy—for example, when the statistical discrepancy in the U.S. balance of payments in the early 1990s signaled that something might be amiss with the data. Are there additional good examples about the circumstances in which the studies proved their usefulness?
4. Are there rules of thumb related to any of the approaches—such as a rule about the size of a statistical discrepancy relative to an estimate? Are these rules of thumb useful?
5. At least one approach—revision studies—may be less applicable for datasets that are subject to few (or no) revisions. Are the other three approaches applicable for other datasets? Are there special features of these that may be exploited further for other datasets?
6. Datasets based on a single survey can turn to statistical methodology, as embodied in a substantial literature, to assess accuracy. Are there other approaches that are more applicable to datasets based on administrative records?

V. POSTSCRIPT

This note has focused on assessment and validation of output data. Assessment and validation are also carried out before the output stage, at the stages in the statistical process that transform source data into intermediate data and finally into output data. A note about assessment and validation at these earlier stages in the process would have an entirely different character, however, because the techniques and approaches tend to be less formal and are internal to the statistical agency; as such, they are much less publicly documented.

In the authors' experience, at least two of the approaches described in this note—statistical discrepancies and comparisons of like data—are used extensively at the source data and intermediate data stages. (The other two approaches—revision studies and judgmental evaluations—by their nature are less amenable at the earlier stages.) A few examples serve to illustrate this point.

- **Statistical discrepancies:** At various stages in the statistical process, two totals that should be equal can be struck and the trial difference can be used to inform next steps. The steps may include further data checking and/or adjustments to one total or both totals. Indeed, it was noted above in Section III. A. that BEA makes adjustments to limit the size of the published statistical discrepancy as it prepares the GDP estimates.
- **Comparison of like estimates:** As noted in Section III. B., this process is familiar to statisticians because it is carried out not only for output data but also for source data and intermediate data. As ready examples, one can think of comparison of customs data and data from international transactions reporting systems (that is systems that capture cash transactions that pass through banks), in countries where both exist, for the goods components of balance of payments. For national accounts, comparisons can sometimes be made of results from two surveys or from an administrative source and a survey. For example, the results of household surveys on expenditures can be compared with retail sales survey data for components of household consumption, or administrative data on wages and salaries from an unemployment insurance program can be compared with data from an enterprise survey of employment, hours, and earnings.