Distr.
GENERAL

CES/AC.68/2004/19
21 April 2004

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD)

CONFERENCE OF EUROPEAN STATISTICIANS

COMMISSION OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat/OECD Meeting on National Accounts
(Geneva, 28-30 April 2004)

MEASURING REVISIONS OF STATSTICAL OUTPUTS

Paper submitted by European Central Bank[1]

SUMMARY

This note suggests the use of indicators for revision studies that allow distinguishing between systematic revisions and unsystematic revisions. This distinction is critical because systematic revisions can be assessed and should, in principle, be integrated in the estimation procedures for earlier estimates.

JEL Classification: C00.
Keywords: Quality indicators; revision studies; forecast error analysis.

INTRODUCTION

1. Quality is a subjective concept and encompasses all aspects of how well statistics meet users' needs. It is intrinsically a multidimensional concept. When applied to statistics, quality is linked to users' expectations about the information content of the disseminated data. For example, data delivered in a timely manner are revised with the aim of fostering economic and financial analyses and reducing noise and uncertainties in modelling, in particular forecasting exercises.
2. This note is based on two references: the cascading structure offered by the IMF Data Quality Assessment Framework (DQAF) for balance of payments (b.o.p.) statistics, which has been

[1] Paper prepared by Jorge Diz Dias. The views expressed in this publication are those of the authors and not necessarily those of the ECB. No responsibility for them should be attributed to the ECB or to any of the other institutions with which the authors are affiliated. Reproduction is permitted provided that the source is acknowledged.

used also as reference background of quality, and the main findings of the Task Force on Output Quality (TF-QA) for b.o.p. and international investment position (i.i.p.), which was co-ordinated by the ECB (DG-Statistics).[2]

3. In the accuracy and reliability dimension of DQAF, the indicator revisions studies is defined as “revisions, as a gauge of reliability, are tracked and mined for the information they may provide”. Different measures are used by statistical agencies to measure this concept of reliability defined as “the closeness of the initial estimated value(s) to the subsequent estimated values”. In a IMF survey about reliability[3] it is stated that “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”.

4. The US Bureau of Economic Analysis (BEA) identifies six sources of revisions:

i) replacement of preliminary source data with revised or more comprehensive data;
ii) replacement of judgmental projections with source data;
iii) changes in definitions or estimating procedures (definitional revisions);
iv) updating seasonal adjustments factors;
v) for real estimates, changes in the base year and changes in the index-number formulas used to calculate them;
vi) and, correction of errors in source data or computations.

5. The purpose of this note is to measure revisions of statistical outputs, using some quality indicators that have been selected to be operational for b.o.p. and i.i.p. outputs at the European level. Measuring quality implies that the various trade-offs among the numerous dimensions and/or elements of quality are taken into account, e.g. between accuracy and timeliness.[4] Against this background, the intention of this note is not to present a unique measure to indicate the overall level of quality of a statistical output, but to identify the different aspects of data quality that could be enhanced.

6. Any indicator for revisions studies should be analysed in the specific context of each statistical output, e.g. if there is a definitional revision then the indicator readings should take this into account. However, it should be noted that, as much as possible, the estimation procedures should anticipate predictable revisions.

7. The structure of this paper is as follows. Section I contains a survey on measures used in revisions studies. Section II describes the reasons why revisions studies indicators should allow identifying systematic and unsystematic revisions. Section III presents the proposed indicators developed by the TF-QA for b.o.p. and i.i.p. to address the issue raised in Section II, keeping the indicators as simple as possible. The conclusion is provided in section IV.

RELATED PUBLICATIONS OF STATISTICAL AGENCIES

[2] Task force mandated by the CMFB whose participants were representatives from Banca d´Italia, Banco de España, Banco de Portugal, Bank of Greece, Banque de France, Central Statistics Office of Ireland, Deutsche Bundesbank, De Nederlandsche Bank, ECB (DG-Statistics), Commission (Eurostat), Office for National Statistics, Sveriges Riksbank, Suomen Pankki and Ufficio Italiano dei Cambi.
[3] Carson, Carol S. and Lucie Laliberté (February 2002).
[4] Timeliness is measured as the length of time beween the end of a reference period and the official release data.

8. This section presents the measures used in public papers about revision studies made by statistical agencies. Seven papers are analysed: one is all about b.o.p. (Australian Bureau of Statistics) and the rest are about National Accounts (four from U.S. Bureau of Economic Analysis and two from U.K. Office of National Statistics). In all these studies a common hypothesis is assumed: later estimates are assumed to be a better approximation of the “true value” because they incorporate more information.

US Bureau of Economic Analysis (BEA)

9. The BEA has an extraordinary experience of revision studies applied to the GDP and GNP (and their components) estimates. In the early papers the BEA used six summary measures for reliability:

Measure	**Expression**
1. Dispersion	*S|E-L| / n*
2. Relative dispersion	*(S|E-L|) / S|L|*
3. Bias	*S(E-L) / n*
4. Relative bias	*S(E-L) / S(L)*
5. Upward revisions	*(# upward revisions) / n*
6. Directional misses	*#(sign E <> sign L) / n*

where *E* stands for earlier estimates, *L* for the later estimates available and *n* for the number observations of the time-series.

10. In the late 90's, BEA revision studies focuses on the first three measures presented on the last table. However, the analysis started to shed light on the correct indication (information) provided by estimates, namely:

Measure	**Expression**
1. Correct indication of change	*#(sign E = sign L) / n*
2. Correct indication of de/acceleration	*#(sign DE = sign DL) / n*
3. Correct indication of the growth trend rate	*#(1.5%£ E £4% and 1.5%£ L £4%) / n*
4. Correct indication of turning points	*#(turning point for E = turning point for L) / n*

11. In the recent article “Reliability of GDP and related NIPA estimates” the previous measures were redefined and additional measures were added:

	Measure	Expression
Revisions distribution	1. Mean absolute revision (MAR)	$S\|L\text{-}E\| / n$
	2. Mean revision (MR)	$S(L\text{-}E) / n$
Deviations distribution	3. Standard deviation of the mean absolute revision (SD)	$\sqrt{\Sigma(\|L-E\|-MAR)^2 / \mathrm{n}}$
	4. Mean absolute deviation (MAD)	$\Sigma \|\| L-E \| - MAR \| / \mathrm{n}$
	5. Coefficient of variation of the absolute revisions (CV)	SD / MAR
	6. Correlation between different vintages peer revisions	$\mathrm{Cov}(R_i; R_{i+n}) \big/ \mathrm{Std}(R_i).\mathrm{Std}(R_{i+z})$

where $R = L - E$, i stands for the vintage, z the number of vintages ahead, $Cov(a; b)$ the covariance between a and b, $Std(b)$ the standard deviation of b.

UK Office for National Statistics (ONS)

12. In the paper "Revisions analysis of initial estimates of annual constant price GDP and its components" reliability "of a series refers to the consistency between the initial estimate and the final estimate of that series". The measures used by this paper and the paper "Revisions analysis of initial estimates of key economics indicators and GDP components" are:

Measure	Expression
1. Mean revision (x)	*Average (R)*
2. Mean absolute revision	*Average* $\|R\|$
3. Standard deviation of revisions (s)	*Average* $(R\text{-}x)^2$
4. Coefficient of serial correlation (α)	$\mathrm{Cov}(R_i; R_{i\text{-}1}) \big/ s^2(R_i)$
5. Standard error of Mean revision (s_x)	$\sqrt{s^2(1+a) \big/ n(1-a)}$
6. t-value	$[x / s_x]$ *with* $n^* = [n(1\text{-}a^2) / (1+a^2)]$ *degrees of freedom if the null hypothesis, that the Mean revision is zero, is true.*
7. Number of upward and downward revisions	*# (R ³ 0) / n and # (R < 0) / n*
8. Range of revision values	*Min (R) and Max (R)*
9. Mean square error	*Average* $(R)^2 = x^2 + s^2$

where R equals $E_{later\ assessment} - E_{1st\ assessment}$, n the number of observations. ONS modified the standard *t-test* to account for positive serial correlation in the revision series. It is also presented an analysis of bias (using the mean revision) and strength of GDP growth.

Australian Bureau of Statistics (ABS)

13. The paper "Quality of Australian Balance of Payments Statistics" of 1996 is a valuable approach to quality[5] assessment in b.o.p.. The reliability is defined as "the proximity of initial and intermediate estimates for a particular period to the final estimate for that period". The measures used are:

Measure	Expression
1. Bias (unscaled)	*Median (E-F)*
2. Bias (scaled)	*Median [(E-F) / F]*
3. Dispersion (unscaled)	*Median |E-F|*
4. Dispersion (scaled)	*Median |(E-F) / F|*
5. First and third quartiles of bias (dispersion assessment)	*Quartiles [(E-F) / F]*
6. Direction of one-year revisions	Unscaled: Median $(E_{1\ year\ later} - E_{1st\ assessment})$; Scaled: Median $[(E_{1\ year\ later} - E_{1st\ assessment}) / E_{1st\ assessment}]$
7. Magnitude (dispersion) of one-year revisions	Unscaled: Median $\lvert E_{1\ year\ later} - E_{1st\ assessment} \rvert$; Scaled: Median $\lvert (E_{1\ year\ later} - E_{1st\ assessment}) / E_{1st\ assessment} \rvert$
8. Sign of revisions	*#(sign (F - E) < 0); #(sign (F - E) = 0); #(sign (F - E) > 0)*
9. *Error* reduction (assuming that F is more accurate)	$\#[(E - E_{1st\ assessment})/(F - E_{1st\ assessment}) \leq 1]$; $\#[(E - E_{1st\ assessment})/(F - E_{1st\ assessment}) > 1]$
10. Revisions alter information provide by initial estimate6	$\Delta F / \Delta E_{1st\ assessment}$

where *E* represents earlier estimates and *F* the final estimate available at the moment of analysis.

DESIRABLE PROPERTIES OF THE INDICATORS

14. The analysis of the usual indicators on revisions, as those previously reviewed (e.g. mean revisions and mean absolute revisions), is very focused on revisions in general. Revision studies should allow for distinguishing the systematic revisions from the unsystematic revisions. This classification is critical because systematic revisions contain regular patterns, while unsystematic revisions are connected to unpredicted changes. Therefore, it is useful to have indicators that can be decomposed into a systematic component and an unsystematic component. By definition, the systematic component can be largely foreseeable and hence tackled by statisticians in advance.

15. Moreover, the indicators should be calculated over series with a minimised deterministic component, i.e. limited trend and cyclical parts. It is important that the deterministic component is removed, using, for example, first differences, growth rate or other method. This part of time series

[5] "The quality of statistics about a topic should be judged in terms of their ability to meet users's needs in that area. To be of high quality, statistics should be: relevant to user's needs (i.e. measure the concepts in which users are interested); accurate; reliable (i.e. no subject to large revisions); comprehensive in coverage; consistent with related statistics; timely and easily accessible". McLennan, W. (February 1996).

6 "Because of the preliminary nature of statistics for the latest months or quarters shown in balance of payments publications, care is needed in interpreting period-on-period (month-on-month or quarter-on-quarter) changes in the magnitude of balance of payments aggregates for the more recent periods". McLennan, W. (1996).

usually follows a very stable pattern. The non-deterministic component of the series should in principle be stationary, i.e. have a constant (unconditional) mean and variance.

16. In addition, the indicators should penalise less series with a higher volatility because they have more associated uncertainty than series with low volatility. Hence, the indicators should allow for higher total revisions in more volatile series since their non-deterministic component is more relevant.

17. The indicators should be constructed in a manner that allows comparability across different series and, if possible, should have no unit.

THE CONTRIBUTION BY THE TF-QA IN THE CONTEXT OF B.O.P. AND I.I.P.

18. Following the literature on measuring the performance of forecasts, one could devise indicators to detect possible bias or persistent patterns in the revisions. The most appropriate one for the analysis of the revisions has been built based on mean square errors (MSE), as it can be decomposed into bias and variance.

19. The use of this measure is based on the following interpretation of the revisions. The early assessment *(E)* is considered the best forecast of the series, estimated with the information available at that moment. The later assessment *(L)* is assumed as the most accurate estimation and thus the closest to the real observation. It is quite useful to perform this analysis for the various vintages of estimates. Consequently, the revision is considered a forecast error, and the indicator based on MSE seeks to measure the quality of the forecast.

20. The indicator which was finally considered the most appropriate by the TF-QA is a ratio between two different mean square errors, making it a relative measure:

- the numerator uses the MSE applied to the difference between two assessments (*revision variation*):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(E_t - L_t)^2 \qquad (1)$$

- the denominator uses the MSE applied to the difference between variable L and a reference value Θ for the variable L (*reference variation*):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(\Theta_t - L_t)^2 \qquad (2)$$

21. This indicator is a relative measure of the revisions expressed as a percentage of reference variation. Usually, the square root is applied to this ratio and is named the root mean square relative error (RMSRE):

$$RMSRE = \sqrt{\frac{\sum_{t=1}^{N}(E_t - L_t)^2}{\sum_{t=1}^{N}(\Theta_t - L_t)^2}} \qquad (3)$$

22. The RMSRE's value is 0 when there are no revisions, 1 if the early estimates are only as accurate as the reference value, and greater than 1 when the early estimates are less accurate than the reference value.

Special case for the reference value

23. If the reference value for L is the constant forecast model of the mean, then the equation (2) becomes the variance of L. The advantage of using the average is that the MSE can be decomposed[7] into three components, with interesting applications for the study of the revisions:

$$MSE = (1)\ Bias\ component + (2)\ Regression\ component + (3)\ Disturbance\ component$$

24. Applying this decomposition to the mean square relative error (MSRE) results in:[8]

$$MSRE = \left[\frac{\overline{E} - \overline{L}}{S_L}\right]^2 + \left[r_{EL} - \frac{S_E}{S_L}\right]^2 + \left[1 - (r_{EL})^2\right] \qquad (4)$$

where E is the early assessment (*forecast)*, L is the later assessment (observed event), r_{EL} is the correlation between the two series, s_E and s_L are the standard deviation of E and L, respectively, and $\overline{E}$ and $\overline{L}$ the average of E and L, respectively.

25. The interpretation of each component is as follows:

- The unconditional or bias component is an indication of systematic error in revisions, since it measures the extent to which the average values of the early and later assessment series deviate from each other. The revisions can be considered biased if the mean revision is significantly different from zero[9].
- The conditional or regression component is another systematic component which reflects whether the overall pattern of the series with the early estimates was close to that of the series with the later estimates. If the early estimate correctly reflects the pattern/variability of the later estimate series, the correlation between both series will be quite high and this component will tend to zero.
- The unsystematic or disturbance component is the variance of the residuals obtained by regressing the early estimates data on the later estimates. It can be assumed to have a random nature and without any predictable pattern[10]. This component truly reflects the improvements in estimates.

Analysis of vintages

26. Together with the calculation of the RMSRE and its decomposition for each data revision vintage, this indicator may be adapted to study the speed of convergence to the final data using in the reference variation the total variation between first (P) and final (F) estimates. Therefore the

---

[7] See Theil, H. (1996) and Stewart, T.R. *et al* (1994).

[8] These three components can be presented as proportions adding to 100%.

[9] Normality is assumed for revisions in order to apply the *t-test*.

[10] This indicator only accounts for linear relationships. The unsystematic part could still have non-linear patterns within it.

RMSRE would take the following expression:

$$RMSRE_{E,L} = \sqrt{\frac{\sum_{t=1}^{N}(E_t - L_t)^2}{\sum_{t=1}^{N}(P_t - F_t)^2}} \qquad (5)$$

where E can coincide with P and L with F.

27. This indicator version gives the *total variation* of a specific revision vintage as a percentage of the *total variation* in revisions. The analysis of $RMSRE_{E,L}$ for consecutive revision vintage provide a good idea on how many vintages are need to reach the final estimate.

CONCLUSION

28. This note stressed the importance that revision studies indicators should allow for distinguishing between the systematic revisions and unsystematic revisions. This decomposition is of most importance because systematic revisions can be statistically tacked, and largely represent foreseeable changes to earlier estimates. The classical measures of revisions, for example mean revisions, do not allow for this breakdown and thus they should be used as a starting point for a deeper analysis of revisions.

REFERENCES

International Monetary Fund

Data Quality Assessment Framework for Balance of Payments Statistics (July 2001 Vintage). IMF - Statistics Department.

Carson, Carol S. and Laliberté, Lucie (February 2002). *Assessing accuracy and reliability: a note based on approaches used in national account and balance of payments statistics*. Working paper IMF - Statistics Department.

European Central Bank

Task Force on Quality - Interim report on the quality assessment of balance of payments and international investment position statistics (November 2003). <http://www.cmfb.org/pdf/TF-QAreport_final_CMFB.pdf>

Bureau of Economic Analysis

Young, Allan H. (Carol S. Carson, Frank de Leeuw and Robert P. Parker contributed to this article) (August 1987). *Evaluation of the GNP Estimates*.

Young, Allan H. (October 1993). *Reliability and accuracy of the quarterly estimates of GDP*.

Grimm, Bruce T. and Robert P. Parker (December 1998). *Reliability of the quarterly and annual estimates of GDP an Gross Domestic Income*.

Fixler, Dennis J. and Bruce T. Grimm (January 2002). *Reliability of GDP and related NIPA estimates*.

Australian Bureau of Statisitcs

McLennan, W. (February 1996). *Quality of Australian Balance of Payments Statistics*.

Office for National Statistics

Barklem, Adèle. 'Economic Trends' No. 556 (March 2000). *Revisions analysis of initial estimates of key economics indicators and GDP components*.

Symons, Peter. 'Economic Trends' No. 568 (March 2001). *Revisions analysis of initial estimates of annual constant price GDP and its components*.

Others

Stewart, T. R. and Lusk, C. M. (1994). “Seven components of judgmental forecasting skill: Implications for research and the improvement of forecasts”. Journal of Forecasting, 13, 579-599.

Thiel, H. (1966), *Applied Economic Forecasting*. Chicago: Rand McNally.

Verbeek, Jos (December 1998). *The World Bank’s unified survey projections, How accurate are they? an ex-post evaluation of US91-US97*. World Bank

* * *