18 □ 2001

WORKING DOCUMENTS

Quality in Short-term statistics: The Euro-SICS database of Eurostat

1

EUROPEAN COMMISSION

eurostat

THEME 1
General statistics

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2001

Printed in Belgium

PRINTED ON WHITE CHLORINE-FREE PAPER

INDEX

Quality in Short-term statistics: The Euro-SICS database of Eurostat

Dominique LADIRAY, Fabio SARTORI
EUROSTAT, Unit A6
Statistical Indicators for Euro-zone Business Cycle Analysis
Jean Monnet Building, L-2920 Luxembourg
e-mail: dominique.ladiray@cec.eu.int, fabio.sartori @cec.eu.int

1. Introduction

Unlike most of the literature about quality in official statistics, which mainly focus on surveys, this paper raises a new issue concerning the measurement of data quality in time series statistics. Our study describes the main features of the Euro-SICS[1] database of Eurostat, which presents a collection of monthly and quarterly business cycle indicators for the Euro-zone, the European Union and the EU Member States.
The database was created to provide data for business cycle analysis and satisfy the needs of the European Central Bank for timely and long time series helpful for devising an appropriate monetary policy for the Euro-zone.
The state of affairs of the database is evaluated monthly through the Eurostat Qualistat approach. Quality indicators, by country and statistical domain, have been produced to measure the relevance, timeliness, completeness, accuracy and clarity of the time series. As there is a demand for long time series, the relevance of the statistics provided by the database has been measured in terms of time series length. Moreover, short-term data are relevant if they are timely. Hence the key quality aspect is timeliness. Data clarity is another important quality aspect and supposed to be achieved through the provision of metadata describing the features of the EU aggregates according to the Special Data

[1] Euro-**SICS** stands for Euro **S**tatistical **I**ndicators **C**ommon **S**ite. Euro-SICS is thus a database of the European Statistical System as a whole to which National Statistical Institutes (NSIs) and National Central Banks (NCBs) are contributing.

Dissemination Standard of the IMF.
A special issue to deal with while measuring the quality of time series is the establishment of quality objectives for time series that not only concern different statistical domains, but whose indicators also have different frequency. Another issue that should be stressed is the accuracy measurement. Analysis of revisions can be carried out to assess the accuracy of estimations. This implies that all the statistical domains and countries concerned have to apply comparable revision policies and all the Euro-SICS domains should adopt the same flagging system so that the different revisions of data can be tracked.
Finally, the main point to be borne in mind is that quality is a dynamic concept, so what today is deemed to be an appropriate measurement of a phenomenon, may not be suitable for future changes in user's needs as well as in the configuration of the database. A continuous monitoring of the former and periodic checks of the latter will let us fulfil our purposes.
The outline of the paper is as follows: Section 2 presents the general features of the Euro-SICS database, focussing mainly on its structure, its users and the provision of metadata. Section 3 discusses some topics regarding quality of official statistics and its evaluation. Section 4 presents some results concerning some indicators of relevance and completeness, such as the length of the series, the number of missing values, etc. Sections 5 and 6 focus, respectively on the evaluation of timeliness and accuracy.

2. The Euro-SICS database

Euro-SICS supplies institutional users with a wide range of short-term indicators for the Euro-zone, the European Union and EU Member States. It is an extension of the Euro-indicators template accessible through Eurostat's web site. The project was launched at the beginning of 1999 and the beta version of Euro-SICS was presented in July 1999. The Euro-SICS database was opened to privileged users at the end of 1999 and a revised version will be accessible to the public at large in June 2001.

2.1 General description of the Euro-SICS database

The Euro-SICS series were selected on the basis of economic and statistical criteria. They have to give an encompassing and fairly accurate picture of the short-term economic situation. Some series were chosen because of their cyclical features (leading-lagging structure), others because of their relevance in the construction of other macroeconomic aggregates. As quarterly national accounts are at the centre of this exercise, mainly those short-term series were retained whose evolution is closely related to the evolution of quarterly account aggregates.
The Euro-SICS series can be classified into three main categories: fully harmonised series; not fully harmonised national series available in all Member States; and Member State specific series. Eurostat is in charge of series from the first two categories while Member States for series from the last category. The Euro-SICS database contains 56 main indicators. With the different breakdowns (by sector, by activity, etc.) the total number increases to approximately 450 indicators. The indicators are organised in 15 domains (see Table **1**)[2].

[2] The classification plan of Euro-SICS has some weaknesses: lack of domain homogeneity and a certain degree of

To assure its usefulness for short-term analysts and policy-makers several principles have to be respected: real-time updating (a delay of one working day between national release and the update of the database); series of sufficient length; and full documentation. Moreover, in view of ever changing needs the database has to be assessed and improved continuously. A working group composed of contributors and key users has been established to monitor the content of the database and the user needs. After the opening of the site to the general public, periodic surveys will be conducted to monitor the satisfaction of users. Eurostat issues a "Monitoring Report" to inform all contributors and main users about the evolution of the content of the database. This report is supplemented by a "Global Quality Index" meant to summaries and condense the quality features of the database in a single synthetic indicator.

2.2 Euro-SICS Metadata

Eurostat adopted in 1999 the SDDS (Special Data Dissemination Standard) for the Euro-SICS database for a host of reasons. The heterogeneity of the database with its many contributing institutions required the adoption of a standard to simplify the drafting and updating metadata files within Eurostat and beyond. The IMF standard had the advantage to exist already, so that Eurostat refrained from developing an own standard despite the shortcomings of the IMF standard. Moreover, as many NSIs and NCBs are already using SDDS to deliver metadata to the IMF, the adoption of this standard made it easier to integrate metadata for national and European series. Finally the adoption of a well-known standard was supposed to be also easier for users.
However, to facilitate SDDS usage, Eurostat has opted for a split of its metadata into three parts: a first one containing only basic information and dissemination formats, a second one containing a summary methodology, and a third one containing the release calendar. The first part gives administrative details about the data (coverage, periodicity and timeliness), about the access for the public, about the integrity (terms and conditions, internal access, ministerial commentary, information about revision and changes in methodology), and about the quality (documentation about the methodology and sources, information supporting cross-checks). A table of dissemination formats is also provided.
The second part on the summary methodology consists of six sections: (1) the analytical framework, concepts, definitions and classifications; (2) the scope of the data (coverage of institutional units, geographic areas, etc.); (3) accounting conventions (time of recording, valuation methods); (4) the nature of the basic data; (5) compilation practices (weighting schemes, imputation methods, etc.); and (6) other aspects such as seasonal adjustments, reference years, etc. The third part (Release Calendar) is directly available from the Eurostat web-site. National release calendars will become available through links with national web-sites.

3. The necessity of a quality measure

Eurostat is, of course, not really a newcomer in infra-annual economic statistics. However, its service was not so well-known and thus not much used in the past. This had to change

redundancy in the classification structure). A new structure has been adopted for the database and will be used for the public release.

and changed also quite rapidly, once the third phase of EMU started. Therefore Eurostat had to improve the notoriety and appreciation of its service by the public at large and, most notably, professional users. Quality measures for Eurostat's infra-annual economic statistics were considered to be of help in this context. For this it was possible to draw upon Eurostat's quality initiative launched in 1998. Within the framework of this initiative (launched also in several Member States) Eurostat developed its (user oriented) quality concept (Eurostat 2000a and 2000b) built on the ISO 8402 norm. This norm states that: "Quality is the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs".

The features and characteristics retained for Eurostat's quality concept are: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, coherence and completeness. However, the definitions seem to have been developed mainly in the framework of survey statistics and their extension to time series statistics is not so straightforward (at least for some of them). Nevertheless they have served quite well as a starting point for assessing the quality of the Euro-SICS database.

All these features and characteristics are, of course, not of equal importance and not all of them are equally relevant in the context of time series statistics. The aspect of timeliness and punctuality is obvious of crucial importance for the Euro-SICS database, the more so as Eurostat has often been criticised for the lack of timeliness of its indicators. Therefore this aspect has to be particularly elaborated in the quality assessment of Euro-SICS.

Moreover, an increasing range of additionally also changing user needs have to be considered. Euro-SICS is for the moment accessible only for a number privileged (and highly professional and knowledgeable) users (NSIs, ECB, ministries, etc.). Once the database will be accessible by the public at large features and characteristics of minor importance in the past (e.g. accessibility, clarity, etc.) will become much more important and have thus to get more attention. A quality assessment such as Eurostat's that is geared towards user needs will have to change when the composition of users is changing.

Finally, however complex a theoretical framework for the assessment of the quality of a database in view of ever changing user needs might be, in the end simple routines have to be developed that can be handled easily. In the case of the Euro-SICS database a couple of fairly straightforward data-mining routines have been set up. These routines, however limited they are, try to cast the theoretical features into simple and still useful numbers. Of course, they are only the starting point, more sophisticated routines have to be developed, most notably also when the underlying database becomes more sophisticated.

4. Relevance and completeness

In Eurostat 2000a, the aspect of relevance and completeness is defined as follows:

Relevance of statistical concepts: The statistical concepts used for statistics are relevant if they meet user needs. The identification of users and their expectations is therefore necessary.

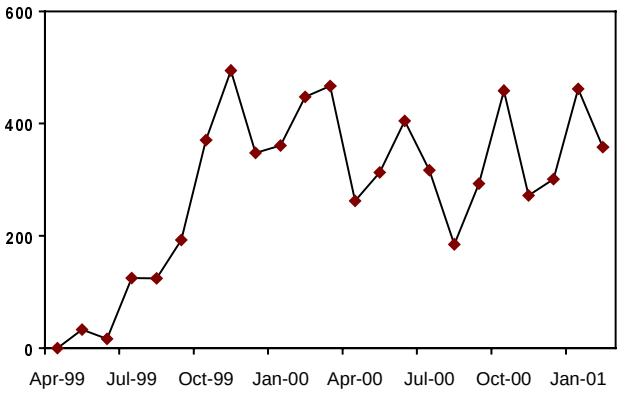
Completeness: Domains for which statistics are available should reflect the needs and priorities expressed by the users of the European Statistical System.

Such a definition of quality with respect to relevance and completeness implies a scrutiny of the database content by users. Therefore a survey has been conducted recently to know more about the needs of Euro-SICS users (a very limited group of NCBs, NSIs and some other institutions such as DG ECFIN, ECB, etc.). Eurostat intends to repeat such surveys

among a wider range of users (e.g. financial market operators) in the future. By the way, the first survey (see Eurostat 2001a) led to a somewhat surprising result, as users wanted "more indicators, but less breakdowns". This stood in stark contrast to both the ECB and DG ECFIN, because these institutions came out in favour of a reduced number of indicators (see ECB 2000). Expressed user needs can obviously be quite contradictory. In this case the official statisticians will have to decide taking into account the resources at their disposal and their technical possibilities.
There are, however, some aspects other than (expressed, assumed or otherwise identified) user needs which allow to assess relevance and completeness. Firstly the length of the time series has a strong bearing on the relevance of a statistical database, as sufficiently long time series are needed for short-term economic analysis (at least two cycles). Moreover the periodicity plays equally an important role, because a short-term analysis can only be conducted on the basis of high frequency data (monthly or quarterly). These aspects are analysed in the following sections.

4.1 Relevance and consultation

Graph 1 provides the number of monthly extractions from the Euro-SICS database ever since the database became available. This indicator gives an idea of how often the database is consulted, and thus of its relevance.



Graph 1: Evolution of the number of extractions from the Euro-SICS database.

It should be noted, however, that some privileged users (ECB and its ESCB partners) can get access to this set of infra-annual economic data not only through Euro-SICS but also through other means: through a direct data delivery by Eurostat or through CIRCA. Both non-standard ways of accessing data have not been taken into account, so Graph 1 underestimates the relevance of the database.

4.2 Coverage and regular updating

The database contains only 56 indicators. However, taking into account regional, sectoral and other break-downs the database contains a theoretical number of approximately 15,000

time-series. For a host of reasons (confidentiality, material or logical impossibility, etc.), not all the Member States are able to compile all indicators.
Table **1** shows the distribution between active series that are regularly updated, non-active series that have not been updated during the last twelve months, and simply missing series that are non-compiled by Member States or Eurostat for whatever reason. The database currently contains 8472 "active" series, while 6119 series are "missing" and 631 series are "inactive".

Series	Missing		Active		Inactive		Total
Domain	N	%	N	%	N	%	N
Balance of Payments	630	57.1	474	42.9	.	.	1104
Business and Consumer Surveys	40	6.5	572	93.5	.	.	612
Energy	.	.	80	94.1	5	5.9	85
External Trade	28	7.4	332	87.8	18	4.8	378
Industry	409	17.6	1774	76.2	146	6.3	2329
Labour Cost	17	12.5	104	76.5	15	11.0	136
Employment, Construction, M	.	.	16	100.0	.	.	16
Employment, Construction, Q	.	.	26	86.7	4	13.3	30
Employment, Industry, M	26	5.0	476	91.2	20	3.8	522
Employment, Industry, Q	143	14.5	668	67.7	175	17.7	986
Monetary and Financial Indicators	84	38.9	66	30.6	66	30.6	216
National Accounts	4432	62.5	2629	37.1	31	0.4	7092
Prices	269	23.6	753	66.1	118	10.4	1140
Retail sales Index	41	10.0	342	83.8	25	6.1	408
Unemployment	.	.	160	95.2	8	4.8	168
Total	6119	40.2	8472	55.7	631	4.1	15222

Table 1: Distribution of the series by length in the Euro-SICS database (by domain as of 01/03/2001, only active series are considered)

4.3 Frequency

The database consists of fifteen domains of which ten contain monthly and the remaining five contain quarterly series. There is no domain with any annual data. In terms of series, roughly 55% of the series are monthly and the remaining 45% quarterly, whereby the domains themselves are homogeneous[3].

4.4 Length

The database is supposed to contain series spanning over at least 15 years. 52.3% of the active series of the database cover already 15 years or more, and 73% cover 10 years at least. This is fairly satisfactory (Table 2). But the situation changes heavily from one domain to the other.
For example, as far as Balance of Payments is concerned, more than 88% of the series cover more than 15 years: on the other hand, 76% of the series on Labour Cost cover less than 10 years. **Error! Reference source not found.** compares the different domains.

[3] This is going to change with the introduction of a new classification plan. Homogeneity of domains with respect to frequency is not an attribute of importance. More important is a classification plan which does not give rise to too many non-existent series.

Length of the series	from 1 year to less than 5 years		from 5 years to less than 10 years		from 10 years to less than 15 years		15 years and more		Total
Domain	N	%	N	%	N	%	N	%	N
Balance of Payments	42	8.9	6	1.3	6	1.3	420	88.6	474
Business and Consumer Surveys	21	3.7	55	9.6	101	17.7	395	69.1	572
Energy	.	.	.	.	23	28.8	57	71.3	80
External Trade	36	10.8	28	8.4	.	.	268	80.7	332
Industry	79	4.5	365	20.6	525	29.6	805	45.4	1774
Labour Cost	.	.	79	76.0	11	10.6	14	13.5	104
Employment, Construction, M	.	.	6	37.5	4	25.0	6	37.5	16
Employment, Construction, Q	.	.	6	23.1	10	38.5	10	38.5	26
Employment, Industry, M	38	8.0	112	23.5	94	19.7	232	48.7	476
Employment, Industry, Q	37	5.5	192	28.7	183	27.4	256	38.3	668
Monetary & Financial Indicators	3	4.5	23	34.8	20	30.3	20	30.3	66
National Accounts	61	2.3	506	19.2	516	19.6	1546	58.8	2629
Prices	1	0.1	390	51.7	95	12.6	267	35.3	753
Retail sales Index	25	7.3	158	46.2	125	36.5	34	9.9	342
Unemployment	.	.	16	10.0	40	25.0	104	65.0	160
Total	343	4.0	1941	22.9	1753	20.7	4433	52.3	8470

Table 2: Distribution of the series by length in the Euro-SICS database (by domain as of 01/03/2001, only active series are considered)

For European aggregates, however, the situation is not as good. Only 43.4% of the Euro-zone series cover more than 10 years. Table 3 reveals quite clearly the unsatisfactory length of Euro-zone series and the same is obviously valid for EU series.

Length of the series	from 1 year to less than 5 years		from 5 years to less than 10 years		from 10 years to less than 15 years		15 years and more		Total
Domain	N	%	N	%	N	%	N	%	N
Balance of Payments	21	*43.8*	.	.	.	.	27	*56.3*	48
Business and Consumer Surveys	.	.	.	.	3	*8.6*	32	*91.4*	35
Energy	.	.	.	.	4	*100*	.	.	4
External Trade	.	.	.	.	.	.	28	*100*	28
Industry	22	*17.3*	75	*59.1*	22	*17.3*	8	*6.3*	127
Labour Cost	.	.	7	*87.5*	1	*12.5*	.	.	8
Employment, Construction, M	.	.	2	*100*	.	.	.	.	2
Employment, Construction, Q	.	.	2	*100*	.	.	.	.	2
Employment, Industry, M	12	*20.7*	24	*41.4*	22	*37.9*	.	.	58
Employment, Industry, Q	12	*20.7*	24	*41.4*	22	*37.9*	.	.	58
Monetary & Financial Indicators	.	.	5	*38.5*	2	*15.4*	6	*46.2*	13
National Accounts	.	.	22	*100*	.	.	.	.	22
Prices	.	.	37	*56.9*	26	*40*	2	*3.1*	65
Retail sales Index	9	*47.4*	10	*52.6*	.	.	.	.	19
Unemployment	.	.	.	.	16	*100*	.	.	16
Total	76	*15*	208	*41.2*	118	*23.4*	103	*20.4*	505

Table 3: Distribution of the series by length in the Euro-SICS database (by domain as of 01/03/2001, only active series referring to the Euro-zone are considered)

4.5 Missing values

The presence of missing values in a database is obviously an indicator of bad data quality. Series with several missing values are difficult to model, and hence useless for most professional users of the database. For the Euro-SICS database, the situation is very positive, as series with missing values only account for 0.2% of the database and more than 98% of the series do not have missing value at all. Moreover, it should be born in mind that some of the values are even missing for reasons that are quite acceptable (e.g. technical or logical impossibility).

Number of missing values	None		1		From 2 to 10		10 or more		Total
Domain	N	%	N	%	N	%	N	%	N
Balance of Payments	456	96.2	.	.	3	0.6	15	3.2	474
Business and Consumer Surveys	572	100.0	.	.	.	.	.	.	572
Energy	85	100.0	.	.	.	.	.	.	85
External Trade	349	99.7	.	.	1	0.3	.	.	350
Industry	1901	99.0	6	0.3	9	0.5	4	0.2	1920
Labour Cost	115	96.6	.	.	2	1.7	2	1.7	119
Employment, Construction, M	16	100.0	.	.	.	.	.	.	16
Employment, Construction, Q	30	100.0	.	.	.	.	.	.	30
Employment, Industry, M	488	98.4	3	0.6	1	0.2	4	0.8	496
Employment, Industry, Q	831	98.6	8	0.9	.	.	4	0.5	843
Monetary & Financial Indicators	123	93.2	5	3.8	1	0.8	3	2.3	132
National Accounts	2654	99.8	2	0.1	2	0.1	2	0.1	2660
Prices	862	99.0	.	.	.	.	9	1.0	871
Retail sales Index	365	99.5	2	0.5	.	.	.	.	367
Unemployment	160	95.2	.	.	.	.	8	4.8	168
Total	9007	98.9	26	0.3	19	0.2	51	0.6	9103

Table 4: Distribution of the series by number of missing values in the Euro-SICS database (by domain as of 01/03/2001)

5. Timeliness

In Eurostat (2000a), the aspect of timeliness, once again from a user perspective, is defined as follows:
Timeliness and punctuality: Most users want up-to-date figures, which are published frequently and on time at pre-established dates.
Timeliness is, of course, a crucial quality feature of short term statistics. An appropriate evaluation of this aspect for the Euro-SICS database is therefore of utmost importance. However, due to the fairly large number of series contained in this database and above all their heterogeneity (compilation methods applied by Member States, revision policy followed by Member States, estimation of Euro-zone and EU aggregates, etc.) it is quite difficult to get a sufficiently uniform grip on this aspect.
The following section tries to clear the conceptual ground around timeliness, taking into account the time series context. Thereafter the timeliness measurements applied for the Euro-SICS database are presented. Finally the future development in the field will be sketched.

5.1 Concepts and definitions

Different indicators can be considered for measure this particular quality aspect:

Timeliness: the time lag between the reference time point (or the end of the reference period) and the time for publication of the statistics.
Punctuality: the correspondence between the planned / promised date and factual date for dissemination of the statistics.
Freshness: the delay between the date the user consults the database and the reference period of the most up-to-date data.
Indicators a) and b) are suggested in Eurostat (1999a) while indicator c) has been proposed by Grünewald (2000). Timeliness and punctuality are defined from a producer's and freshness from a user's the point of view.
Conceptually at least seven dates can be associated with each national figure in the database:

- the end of the reference period (i.e.: if the reference period is November 99, then d_1 = 30 Nov 99);
- the announced date for its publication (d_2 = 23 Dec 99);
- the date it has been published (d_3 = 23 Dec 99);
- the date it arrives to Eurostat (d_4 = 24 Dec 99);
- the date it appears in Euro-SICS (d_5 = 7 Jan 00);
- the date of the last update of the domain the series belongs to (d_6 = 9 Jan 00);
- the date of the database quality control (d_7 = 15 Jan 00).

For the moment a simple delay indicator is compiled which corresponds to the different $d_6 - d_1$ for the last available data element of the series. It has been defined in this way since the two dates are easily available from Euro-SICS. However, to compile more meaningful timeliness indicators, it is, of course, necessary to monitor the evolution of the Euro-SICS database in conjunction with the release calendars of Member State. Therefore some improvements are planned to obtain a deeper evaluation of all the different aspects concerning timeliness in Euro-SICS.

a) *Timeliness*

Timeliness in a narrow sense could be defined as $d_5 - d_1$. However, this indicator is not easy to compute because Euro-SICS can be updated every day. So a daily check of the database would be necessary to detect new values as soon as they are inserted.

b) *Punctuality*

Punctuality refers directly to a release calendar and pre-defined dissemination dates. If the different release calendars can be linked to the database, it would possible to compute:

- Publication Punctuality = $d_3 - d_2$.
- Euro-SICS Punctuality = $d_5 - d_2$. This delay can be divided in:
- a Member State component = $d_4 - d_2$;
- a Eurostat component = $d_5 - d_4$.

c) *Freshness*

This is the indicator suggested by Grünewald: Freshness = $d_7 - d_1$. There is obviously no problem to compute it, but it depends equally on the timing of the quality control and is thus not a feature related entirely to the data, but partly to the control management.

5.2 The choice of thresholds

These different aspects are essential to assess the quality of the database. They must therefore be part of a global quality index. However, to compile such an index in a meaningful way, it is necessary to define adequately a scale and set a standard for the

various aspects.
As far as punctuality is concerned, adequate standards can be set easily:
there should, of course, be no delay between the date promised in the release calendar and the actual dissemination date, therefore the "Publication Punctuality" standard is 0;
the Member States are supposed to send the data to Eurostat as soon as possible and a realistic objective is that Eurostat receives data the very same day. Eurostat domain managers should also be able to update the database immediately. Their and the standard should also be 0.
For timeliness and freshness, the choice of an appropriate threshold is much more difficult. A threshold should in principle reflect first of all a compromise between the requirements of users and the resources at the disposal of producers. However, other aspects should equally be taken into account. What can realistically be expected by users is, of course, at least what producers elsewhere providing. Thus standards could be set on the basis of a comparison between the EU and the United States[4]. The current performance of the US statisticians could be used as the standard to be matched by EU statisticians. Finally, the statistical domains are quite different (e.g. speed and mode of data collection). Therefore different thresholds should be established for different domains and in some cases even for different series in the same domain.

5.3 Timeliness in the Euro-SICS database

Among the various indicators introduced so far (timeliness, punctuality and freshness), timeliness is for the moment quite difficult to compute since an automatic tracing of the daily modifications of the database has not yet been developed. But even punctuality is difficult to compute, because the database is not yet linked to Member States release calendars. Improvements, however, are planned in near future, so that a more complete picture of the different aspects of timeliness will be available for a more elaborate quality evaluation.
For the moment we can only evaluate the freshness of the database, that is timeliness from the users' point of view. The results of this analysis are presented in Table 5. The results vary considerably across domains. This means that improvement are not required everywhere, but also that different thresholds will probably have to be adopted for different domains (or indicators).

6. Accuracy

In Eurostat 2000a, the following definition is suggested for accuracy:
Accuracy of estimates: Accuracy is defined as the closeness between the estimated value and the (unknown) true population value. Assessing the accuracy of an estimate involves analysing the total error associated with the estimate.
Obviously this definition has been developed and is thus well suited for survey data, but not so much for time series. For time series, the availability of high-quality metadata is probably a very important factor for an accuracy evaluation (see Section 2.2). Nevertheless,

[4] Recently this issue raised a deep debate inside the European Statistical System and lead to the establishment of a Task Force for EU-US comparison, involving Eurostat, the European Central Bank, the OECD and several Member States.

there are also simple ways to measure the accuracy of a time series database such as Euro-SICS. This could be done e.g. by the number of missing values in the series, the number of irregular or "strange" values in the series (outliers), and an analysis of the revision patterns.

Delay of the series	Under 30 days		From 30 Days to less than 60 days		From 60 days to less than 90 days		From 90 days to less than 180 days		Over 180 days		Total
Domain	N	%	N	%	N	%	N	%	N	%	N
Balance of Payments	.	.	.	.	.	.	346	73.0	128	27.0	474
Bus. and Cons. Surveys	572	100	.	.	.	.	.	.	.	.	572
Energy	.	.	.	.	16	18.8	34	40.0	35	41.2	85
External Trade	.	.	186	53.1	146	41.7	.	.	18	5.1	350
Industry	.	.	34	1.8	955	49.7	713	37.1	218	11.4	1920
Labour Cost	.	.	.	.	19	16.0	78	65.5	22	18.5	119
Employment, Const, M	.	.	.	.	.	.	16	100	.	.	16
Employment, Const, Q	.	.	.	.	.	.	16	53.3	14	46.7	30
Employment, Ind., M	.	.	.	.	202	40.7	268	54.0	26	5.2	496
Employment, Ind., Q	.	.	218	25.9	.	.	418	49.6	207	24.6	843
Mon. & Fin. Indicators	30	22.7	.	.	36	27.3	.	.	66	50.0	132
National Accounts	.	.	.	.	776	29.2	1752	65.9	132	5.0	2660
Prices	.	.	459	52.7	278	31.9	5	0.6	129	14.8	871
Retail sales Index	.	.	16	4.4	208	56.7	110	30.0	33	9.0	367
Unemployment	112	66.7	32	19.0	8	4.8	8	4.8	8	4.8	168
Total	714	7.8	945	10.4	2644	29.0	3764	41.3	1036	11.4	9103

Table 5: Distribution of the delay between the reference period and the date of the quality control, that is 01/03/2001 (by domain)

The first approach has already been dealt with above and the third approach cannot be pursued at least for the time being due to the limits imposed by the technical characteristics of the Euro-SICS database. Therefore only the second aspect can be dealt with here.

6.1 Irregular values and outliers

The presence of irregularities in a series is often a good indicator of bad quality. Series containing "strange" values or outliers are generally difficult to model and thus fairly useless. This aspect has been stressed also by database users (see ECB(2000b)). Of course, such an indicator is not supposed to have any "objective" value. It is cannot be anything else than a monitoring indicator that can draw the attention to quite some series and therefore detect possible errors.
To obtain an irregular or "strange" values indicator the database has been screened with the TRAMO software to detect the number of irregular values (outliers) in each series. TRAMO can detect different kinds of "irregularities" in time series, such as additive outliers, level shifts and transitory changes. However, only a (relevant) subset of the database has been analysed for obvious reasons. Inactive series and series with missing values were excluded and so were series with missing values as well as some series that displayed such an irregular pattern that they could not be properly analysed by the TRAMO software. Table 6 shows the results for the different domains of the database. They vary a lot from one domain to another, and they are, at least for some domains, not so satisfactory.

However, as the TRAMO software has been used with the default settings, an *ad hoc* tuning of the parameters may be needed to compute a somewhat better quality indicator.

Number of irregular values	None	1	2	3 and more	4	5 or more	Total
Domain	%	%	%	%	%	%	N
Balance of Payments	31.1	27.4	19.9	9.2	4.6	7.8	412
Business and Consumer Surveys	28.4	23.1	19.1	5.6	7.3	16.5	550
Energy	22.5	35.0	13.8	10.0	7.5	11.3	80
External Trade	18.3	16.3	21.4	14.6	11.5	18.0	295
Industry	24.1	20.3	17.8	12.7	8.1	17.0	1645
Labour Cost	58.4	23.8	13.9	2.0	2.0	.	101
Employment, Construction, M	.	31.3	12.5	6.3	6.3	43.8	16
Employment, Construction, Q	27.3	40.9	9.1	4.5	9.1	9.1	22
Employment, Industry, M	8.0	7.0	11.7	10.3	7.7	55.2	426
Employment, Industry, Q	38.6	24.9	15.2	10.2	6.6	4.6	591
Monetary and Financial Indicators	31.7	20.0	18.3	11.7	1.7	16.7	60
National Accounts	47.2	23.8	14.2	6.9	4.0	4.0	2543
Prices	18.9	15.8	14.5	10.0	10.3	30.4	739
Retail sales Index	41.8	24.1	12.3	8.9	5.4	7.6	316
Unemployment	36.3	22.5	11.3	8.1	5.0	16.9	160
Total	33.0	21.5	15.7	9.2	6.5	14.1	7956

Table 6: Distribution of the series by number of irregular values in the Euro-SICS database (by domain as of 01/03/2001)

6.2 Revisions

Time series of revisions must be tested for any systematic behaviour. As Eurostat considers this to be important a strategy for the analysis of revisions has been described in Eurostat (1999b) for the National Accounts domain. This strategy can, of course, be generalised to other kinds of series beyond the National Accounts domain. The proposed approach is based on the studies of Patterson (1992), Di Fonzo, Pisani and Savio (1994) and many others. Summary statistics can be used to compare between preliminary and final estimates. Two types of errors are distinguished in Eurostat (1999b):

a) relative errors, giving information on the accuracy of preliminary estimates of levels;

b) absolute errors, used to evaluate the accuracy of preliminary growth rates.

For each error type (level and growth rate) a series of accuracy indices for the preliminary estimate can be compiled based on (1) the mean relative error, (2) the mean absolute relative error, (3) the standard deviation of the relative error, (4) the square root of the mean quadratic relative error, and (5) the bias component of the mean quadratic relative error. Clearly, all these measures cannot be considered as reliability measures of provisional estimates. However, they allow to find out if revisions have always (or almost always) the same sign. If this is the case, a correction of preliminary estimates would be needed. The compilation of such revision indicators requires a daily checking of the entire database to capture the complete history of each series. Thus it seems difficult to compile such indicators for all database series . Therefore some further tests are required to select the a group of the series for which this in-depth revision analysis will be performed.

Bibliography

[1] European Central Bank (2000a), *Euro-SICS: proposed coverage and features for implementation by end-2000*, 8 September 2000, note ST/GEFS/WGGES/GES/SICS1

[2] European Central Bank (2000b), *Euro-SICS: comments on Eurostat's Quality Report*, 7 December 2000

[3] Eurostat (1999a), *How to fill in the Quality Report*, Unit A4.

[4] Eurostat (1999b), *Handbook of Quarterly National Accounts*, Unit A4.

[5] Eurostat (2000a), *Definition of Quality in Statistics*, Working Party Assessment of the Quality in Statistics, 4-5 April 2000, Unit A4.

[6] Eurostat (2000b), *Standard Quality Report*, Working Party Assessment of the Quality in Statistics, 4-5 April 2000, Unit A4.

[7] Eurostat (2001a), *Analysis of Main Results of the Euro-SICS users survey*, Euro-SICS Working Party, 15-16 January 2001, Document 047/01, Unit A6.

[8] Eurostat (2001b), *Extension of the List of Country Specific Indicators*, Euro-SICS Working Party, 15-16 January 2001, Document 051/01, Unit A6.

[9] Eurostat (2001c), *Euro-SICS detailed Analysis as at 27 November 2000*, Euro-SICS Working Party, 15-16 January 2001, Document 051/01, Unit A6.

[10] Eurostat (2001d), *A Monitoring Report for Member State Representatives*, Euro-SICS Working Party, 15-16 January 2001, Document 053/01, Unit A6.

[11] Grünewald, W., (2000), *Freshness of NEWCRONOS data: Analysis of the situation as of April 2000*, internal note, Eurostat.