ESA/STAT/AC.202/S2.2

International Seminar on Early Warning and Business Cycle Indicators

14 to 16 December 2009
Scheveningen, The Netherlands

Overview of terminology for high frequency indicators

Roberto Barcellan – Eurostat

International Seminar on Early Warning and Business Cycle Indicators

14 - 16 December 2009, Scheveningen, The Netherlands

Glossary of statistical terms associated to the estimation of key economic indicators

Roberto Barcellan, Gian Luigi Mazzi, Gaetana Montana, Rosa Ruggeri Cannata[1]

Eurostat

Abstract

Statistics are used as a daily reference and source of information to interpret how economic and social phenomena are evolving. The wide dissemination of statistical data and their use in the current language have generated a sort of "statistical jargon" that, often, creates more confusion than clarification. This is inter alia true for the terms used with regard to the estimation of key economic indicators.

The most recent thematic seminars and conferences on the statistical consequences of the financial and economic crisis highlighted even more the need to clarify "what is what" and suggested to set up of a glossary of statistical terms related to the estimation of key economic indicators.

The purpose of this document is to propose such a glossary. The paper has to be considered as a starting point for a discussion in the official statistical community, and not only, towards a meaningful operational terminology in this field.

Key words: glossary, statistical terms, economic indicators, statistics, forecasting.

[1] The authors would like to thank Tomas Duchon (Eurostat) for the input provided in preparing the document.

1. Introduction

Statistics are used as a daily reference and source of information to interpret how economic and social phenomena are evolving. Users require more and more timely statistical information in order to increase their ability to monitor the economic situation and to design effective economic and monetary policies. As a matter of fact, increasing timeliness of regular estimates is not easily achievable in the short-term. For this reason several statistical offices have decided to release more timely estimates by means of statistical end econometric techniques to integrate, complement and in some cases replace, the statistical production process.

Unfortunately, a certain lack of harmonization in labeling all these different statistical estimates has been observed. This is inter alia true for the terms used with regard to the estimations of key economic indicators.

The most recent thematic seminars and conferences on the statistical consequences of the financial and economic crisis have raised the need to clarify "what is what" and to set up a glossary of statistical terms associated to the estimation of key economic indicators.

The purpose of this document is to set up the framework and propose a first draft of a glossary of statistical terms on the estimation of key economic indicators. The paper has to be considered as a starting point for a discussion in the official statistical community towards a meaningful operational terminology on different kinds of estimates of economic indicators.

Section 2 briefly highlights the scope of the glossary, its rationale, its contents, its coverage and the process for its incremental development. Section 3 describes the different dimensions that play a role in qualifying the statistical terms associated to the glossary on key economic indicators, starting from the concepts of statistics and forecasts. Finally, Section 4 outlines some preliminary conclusions. The paper is complemented by two annexes: Annex I illustrates, with examples derived from available statistical glossaries, how the definition of some statistical terms could look like, and Annex II presents more technical details associated to some of these terms providing conceptual insights.

2. Scope of the glossary

- *Why a glossary?*

Several statistical glossaries are available in the literature or on the websites of statistical authorities. Nevertheless, the lack of consensus on the characteristics and boundaries of the statistical concepts covered requires a qualification of the statistical terms associated to the estimation of key economic indicators. A common operating terminology should facilitate the discussion and exchange of opinions among official statisticians and users, fostering a common understanding.

- *Contents*

The approach followed in setting up the glossary can be characterised as "simple and precise". The glossary is intended as a simple list of definitions that clearly identify the statistical concepts within a conceptual framework. Focus is put on short-term key economic indicators.

- *Coverage*

In terms of coverage, the glossary is limited to the statistical terms and concepts related to estimation and to the associated concepts.

- *Process*

The set up of the glossary is an incremental process with a starting point (this document, based on the first exchange of views established in the International Seminar on Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends, held in Ottawa on 27-29 May 2009), followed by discussions in appropriate fora and seminars, notably the International Seminar on Early Warning and Business Cycle Indicators, to be held in Scheveningen, The Netherlands, on 14 – 16 December 2009.

This step-by-step process should ensure the availability of a core set of statistical terms referring to the estimation of short-term economic indicators that potentially will expand in the future on the basis of a built-in consensus in the scientific and official statisticians communities.

- *Insights – technical aspects*

The conceptual framework underlying the selected statistical terms will be progressively complemented by an additional information layer that will analyse more in depth related methodological aspects. This progressive extension to related terms will follow a step-by-step incremental approach.

3. Statistics, nowcasts and forecasts

The starting point for the reflection on the glossary has been the discussion on terminology started in the International Seminar on Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends, held in Ottawa where there was not an unanimous consensus on what terms like statistic, nowcasting, forecasting, flash estimate, leading and coincident indicator stood for.

The target is to identify and qualify the estimation (= assessment) of a measure of a given short-term economic indicator for a given reference period in a specific point in time.

In doing so, the following dimensions have to be considered:

- the relative position in **time** of the estimation with respect to the reference period;
- the **information set** used for the estimation;

- the statistical **method** applied for the estimation;
- the adherence of the estimation process to the regular **production process**;
- the **target variable** to be assessed (original economic indicators vs. proxies).

The cross-interrelations of these dimensions result in a matrix whose cells correspond to the statistical terms and concepts covered by the key terms of the glossary.

Whilst, in principle, the same importance could be attached to all these dimensions, in practice, in defining the different kind of estimates of economic indicators, priority is given to specific dimensions.

In our approach, time plays a prominent role, followed by the information set and then by the method and the adherence to the production process (at the same level) and, finally, the target quantity.

a. The time dimension: statistics vs. nowcasts vs. forecasts –

Let's focus on the estimation of a measure of a given short-term economic indicator *y* for a given reference period *t* in a specific point in time *r*: $y_{(t,r)}$

Then, with respect to time:

- if $t < r$ i.e. we are attempting to estimate a past measure of *y*, then we refer to the estimation as a statistic;
- if $t > r$ i.e. we are attempting to estimate a future measure of *y*, then we refer to the estimation as a forecast;
- if $t \tilde{} r$ i.e. we are attempting to estimate a contemporaneous measure of *y* (just before or after the end of the reference period *t*), then we refer to the estimation as a nowcast.

The estimation (statistic, nowcast or forecast) is further qualified by the other dimensions.

Example 1.a: Eurostat's definition of flash estimate

> **Flash estimate:** An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.a: (a) A nowcast is a contemporaneous estimate of *y* just before or after the end of the reference period *t*.

But this is not enough to qualify a nowcast. A forecast is also covered by this qualification; an early estimate (statistics) too.

b. Information set

A key element to qualify an estimate is the information set used to produce it. The information set corresponds to the input to the estimation process and broadly speaking, in this field of investigation (economic indicators), it corresponds to:

- qualitative information;
- contemporaneous quantitative information;
- past quantitative information.

The qualification of the information used is clearly related to the other dimensions. It might depend on, for example, the moment when the estimation is run, the composition of the information set and its time coverage. Forecasts, nowcasts and statistics are all based on the combination to a greater or lesser extent of the above mention input. The nuances among the different estimates result, therefore, from the different mix of information in a specific point in time.

Example 1.b: Eurostat's definition of flash estimate

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.b: (a) A nowcast is a contemporaneous estimate of *y* just before or after the end of the reference period *t*.

(b) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

But this is not enough to qualify a nowcast. A coincident indicator is also covered by these qualifications.

c. Statistical methods

Different statistical/econometric methods may be applied to produce an estimate (statistics, nowcasts or forecasts). Several families of methods are available for the estimation of key economic indicators. The choice between them depends on the combination of the other qualifying dimensions and on the available techniques and models. The time dimension and the

available information play a prominent role in the choice. The methods range from more official statistical oriented ones (traditional grossing-up, sampling, combination of existing information within the business architecture of statistical authorities) to more statistically oriented techniques (such as extrapolation, structural models, etc.).

Statistical methods, combined with the other qualifying dimensions, determine also the statistical models applied to derive the estimation of the target variable.

It is worth to note that, whilst considerations associated to economic modelling should be avoided in setting up a statistical model and the related statistical method(s), in practice, implicit economic assumptions underline the compilation of official estimations of key economic indicators (e.g., the concept of GDP and its decomposition in components is directly derived from the neo-Keynesian theory).

Example 1.c: Eurostat's definition of flash estimate

> **Flash estimate:** An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.c: (a) A nowcast is a contemporaneous estimate of *y* just before or after the end of the reference period *t*.

(b) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

(c) A nowcast relies on specific dedicated statistical methods that optimise the use of the available information.

But this is not enough to qualify a nowcast. A coincident indicator is also covered by these qualifications.

d. Adherence to the regular production process

In the field of official statistics, particular relevance assumes the regular production process that reflects the business architecture in place in National Statistical Institutes or other statistical authorities. The regular production process corresponds to the "normal" way of producing official statistics on the basis of the available information. In particular, in the case of key economic indicators, it embeds the way of deriving macro-economic aggregates, usually at national level, from the basic information composed by administrative sources, results from surveys, etc., as well as the accounting constraints (if any) generated by the framework of the compilation (e.g. the national accounts framework for GDP).

The regular production process covers also additional characteristics such as revision and release policy that help to further qualify the estimation. Therefore, terms like advanced, preliminary, first, second, regular, final estimate are derived from these additional characteristics.

Example 1.d: Eurostat's definition of flash estimate

> **Flash estimate:** An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.d: (a) A nowcast is a contemporaneous estimate of *y* just before or after the end of the reference period *t*.

(b) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

(c) A nowcast relies on specific dedicated statistical methods that optimise the use of the available information.

(d) The information set and the methods used in producing a nowcast may differ from the methods applied in the regular production process of the target economic indicator.

This qualifies a nowcast.

e. Target variable: indicators vs. proxies

Sometimes the direct estimation of the latest value of the target variable is not feasible or it is technically too complex. Therefore, in order to give an appreciation of the recent or future evolution of the target variable over time, the object of the estimation is not the variable itself but a proxy that well approximates the characteristics and the behaviour of the target variable. In this family, that incorporates for example leading and coincident indicators, the relevant characteristics are the leading-coincident properties of the proxies.

Example 1.e: Eurostat's definition of flash estimate

> **Flash estimate:** An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.e: (e) A coincident indicator is a contemporaneous estimate (nowcast) of a proxy of *y*.

4. Some preliminary conclusions

The task of setting up a glossary on the estimation of key economic indicators is not an easy one. The prominence given to the different dimensions identified in this paper changes the users'/producers' perspective on definitions and terms. This is the reason for which a common

operation terminology, the glossary, is necessary to have a common understanding of "what is what". The paper proposes a building process for setting up the glossary and introduces the key elements (dimensions) to be taken into consideration.

In addition, the paper highlights also the fact that, due to the complexity of the matter, additional and more detailed qualifying elements are necessary for a technical description of the concepts, beyond their pure and simple definition (which remains the main target of the glossary). Therefore, it proposes to complement the concept/definition associated to the terms with more detailed technical considerations ("insights") that explore the complexity of a term. The "insights" resume also natural links with other glossaries of statistical related terms like, for example, a glossary on seasonal adjustment, general statistics glossaries, specific statistical areas glossaries (e.g. national accounts), etc. Annex II gives some preliminary examples of these insights.

References

Information on PEEIs: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-81-08-398/EN/KS-81-08-398-EN.PDF;
http://epp.eurostat.ec.europa.eu/portal/page/portal/euroindicators/peeis

The OECD Glossary of Statistical Terms: http://stats.oecd.org/glossary

BusinessDictionary.com: www.businessdictionary.com

Report on the seminar in Ottawa:
http://unstats.un.org/unsd/nationalaccount/workshops/2009/ottawa/AC188-5.PDF

Annex I: Glossary of statistical terms associated to the estimation of key economic indicators

Coincident indicators: Economic and financial market indicators which tend to move in step with general economic trends such as gross domestic product, employment levels, retail sales, and/or financial market trends such as interest rates and stock market prices.
Synonyms: concurrent indicators
Source: BusinessDictionary.com

Estimate: In the strict sense, the particular value yielded by an estimator in a given set of circumstances.
Source: The OECD Glossary of Statistical Terms

Estimator: A rule or method of estimating a parameter of a population.
Source: The OECD Glossary of Statistical Terms

Extrapolation: Statistical technique of inferring unknown from the known. It attempts to predict future data by relying on historical data, such as estimating the size of a population a few years from now on the basis of current population size and its rate of growth. Extrapolation may be valid where the present circumstances do not indicate any interruption in the long-established past trends. However, a straight line extrapolation (where a short-term trend is believed to continue far in into future) is fraught with risk because some unforeseeable factors almost always intervene.
Source: BusinessDictionary.com

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.
Synonyms: advance estimate (BEA); preliminary estimate (ONS), rapid estimate (Ottawa conference)
Source: Eurostat Handbook on Quarterly National Accounts

Forecasting: and **prediction:** often used synonymously in the customary sense of assessing the magnitude which a quantity will assume at some future point of time: as distinct from "estimation" which attempts to assess the magnitude of an already existent quantity. For example, the final yield of a crop is "forecast" during the growing period but "estimated" at harvest.
Source: The OECD Glossary of Statistical Terms

Hard data: Reliable and methodologically sound data from official or organizational statistics that is comparable and roughly independent from the way it was measured.
Synonyms: factual, quantitative data
Source: Eurostat

Interpolation: Estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information.

Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. Time series analysis and regression analysis are the two statistical techniques employing the concept of interpolation.
Source: BusinessDictionary.com

Nowcasting: A very early estimate produced for an economic variable of interest over the most recent reference period calculated on the basis of incomplete data using a statistical or econometric model different from the one used for regular estimates. Soft data should not play a predominant role in nowcasting models. Nowcasts may be produced during the very same reference period for which the data is produced.
Source: Eurostat

Soft data: Other than hard data, generally less reliable.
Synonyms: qualitative data
Source: Eurostat

Regular estimate: Estimate produced and published for an economic variable of interest over all reference periods normally produced and published using a statistical or econometric model based largely on hard and complete data.
Source: Eurostat

Annex II: Insights – Technical qualifications

A.1. Extrapolation

Strictly speaking in mathematical terms, extrapolation is the process of constructing new data points outside a discrete set of known data points. In other words, extrapolation techniques allow the computation of latest and future values of a time series starting from the already existing values. Extrapolation techniques are purely univariate, so that new values are computed on the basis of the past history. Extrapolation techniques can be either parametric or non parametric, deterministic and stochastic, linear and non linear. Those techniques vary from the simple curve extrapolation (linear, polynomial and logistic, etc.) to the smoothing ones (exponential smoothing, moving average, linear time series models (ARIMA, unobserved components models), and to the non linear time series models (ARCH, GARCH, TAR, STAR, SETAR, etc.). Extrapolation techniques can be applied either directly on the reference variable or indirectly on its components. In the second case the new values of the reference variable will be derived by aggregation of the computed values of the components on the basis of the specific aggregation constraints, reflecting the breakdown structure of the considered variable. Extrapolation techniques can be used before, during and after the reference period and the time horizon can cover 1, 2 or even more periods taking into account that the quality of results will progressively worsens with the length of the horizon. By the use of extrapolation techniques we implicitly assume that the recent past, present and future evolve with the same features of the observed past. This means that the extrapolation techniques can work very well in periods characterized by regular evolution but they will be by definition unable to detect turning points or regimes changes. Despite the limits described above, extrapolation techniques can be very useful when no additional information is available on the reference variable, but there is a need to compute recent or future values. Extrapolation techniques can also be included in more complex rapid estimation processes such as flash estimates to produce new values for geographical or sectoral components of the target variable (possibly of limited weight in the total), for which additional source of information can not be found.

A.2. Nowcasting

In general terms nowcasting is the real-time or quasi real-time estimation of the evolution of a given variable. More specifically, nowcasting allows estimating the latest or the current movements on the reference variable. Nowcasting can take place, right before or after the end of the reference period. For example, if the last available information of the reference variable refers to the period t-1, its nowcasting for the period t should take place a few days before the end of the period t or few days after the beginning of the period t+1. Nowcasting is usually based on an incomplete information set with respect to the one used for the normal compilation process. The information set can be incomplete either from a geographical or a sectoral point of

view, as well as in temporal terms: e.g. for a quarterly reference variable, only two months out of three of the related variables included in the information set are available. Concerning the type of the indicators included in the information set they can be:

- partial information on the reference variable (in geographical, sectoral or temporal way);
- qualitative data of the same sector and nature of the reference variable;
- other quantitative variables, statistically and logically related to the reference variable.

The information set should not contain variables which implicitly indicate the acceptance of a specific economic theory.

Nowcasting techniques can cover a large variety of multivariate methods such us: regression method, multivariate time series methods (e.g. VAR), static and dynamic factor, principal components methods, etc. In order to increase the accuracy of estimates, methods for combining estimates from different statistical methods can also be used. The statistical methods used in the nowcasting context do not necessarily intend to replicate those used in the regular compilation process, but they aim to produce the most reliable results for the reference variable. The main focus on the forecasting exercise is the estimation of the evolution of the reference variable. For these reasons, they are usually applied to the reference variable and their outputs are either growth rates on the previous period or in the corresponding period of the previous year. When nowcasts are published, past values of the reference variable are usually not revised in order to avoid confusion among users. Such confusion could be generated by the differences between the nowcasted values and those coming from the regular compilation process, both on the information set and on the calculation methods. Nowcasting exercises should be carefully assessed, possibly in real-time, before taking a decision on the dissemination. Particular attention should be paid to the absence of a systematic bias between the nowcast and the first regular estimates of the reference variable.

A.3. Flash estimates

Flash estimates aim to provide an early picture of the situation in a particular sector or in the whole economy based on one or some reference variables, possibly related by accounting and or aggregation constraints. Flash estimates will take place after the end of the reference period, based on an incomplete set of information when the amount of available information is considered sufficient to ensure a high degree of compatibility with the first regular calculation. As in the case of nowcasting the set of information can be incomplete in geographical, sectoral or temporal terms, nevertheless in the case of flash estimates the amount of partial information on the reference variable or variables should be higher than in the nowcasting case. In addition the use of qualitative information should be limited and mainly addressed to fill temporal gaps of some variables. The information underlying flash estimates should allow the possibility to replicate, even if in simplified way, the same compilation process used in the regular statistical process. Statistical methods used in compiling flash estimates should be in line with the regular production process. In this context, the statistical methods can vary from simple aggregation techniques to regression methods, to temporal disaggregation and benchmarking ones, to

grossing up techniques applied to small or incomplete sampling scheme, etc. In the case of flash estimates, the use of combining alternative estimates techniques is not considered, unless it is also used in the regular compilation process. Flash estimates will lead to the estimation of level of reference variable or variables even if in the dissemination phase the focus could be more on the period on period and year on year growth rates. Since flash estimates will tend to approximate the same strategy used for regular calculation, they, in principle, may produce a complete time series, with revised values for the previous periods. The opportunity of disseminating such revised figures has to be carefully evaluated case by case. Alternatives are the dissemination of past revised values or just the latest value of the reference variable or variables. Flash estimates should be assessed over a sufficiently long time interval, possible within a real-time exercise in order to analyze their behaviour with respect to the first regular calculation. They could be also benchmarked with the nowcasting of the reference variable and with the result of extrapolation methods in order to compare the relative performances of alternative rapid estimates strategies. Finally, as in the case of nowcasting, particular attention has to be paid to the absences of a systematic bias between the flash estimates and the first regular estimates of the reference variable.

A.4. Coincident indicators and Leading indicators

The approaches discussed in the previous sections aim to estimate the latest missing values for a given time series, which represent the main variable of interest. The methods presented in this section are based on a completely different philosophy. Those methods aim to compile a new time series (indicators) able to reproduce the same movements of the reference variable with a higher timeliness, instead of completing the reference variable with estimates for the latest points. Indicators can be classified into three main categories: coincident indicators, leading indicators, lagging indicators.

Coincident indicators reproduce the movements of the reference variable without temporal shift, so that the movements at time t observed in the indicators correspond to those ones of the reference variable at the same time.

Leading indicators are able to anticipate the movements of the target variable of a fixed number of periods.

Lagging indicators reproduce the movements of the reference variable with some time-lags. Obviously, in the context of this paper only the first two kinds of indicators are relevant. Coincident indicators can be computed during the reference period or right after the end of that. Leading indicators are usually computed before the start of the reference period. Usually, coincident indicators produce one step ahead estimates where for leading indicators it is possible to produce multistep ahead estimates, depending on the characteristics of the data used. Data involved in the construction of the coincident and leading indicators can stem from different areas like survey data, financial data, real sector data, etc. They can also include variables from other economies than the reference one, or measuring relationships among economies. In principle, the amount of data used for the construction of coincident and leading indicators can be very large, up to thousands of series, which is another significant difference with the methods presented in the previous sections. Estimation strategies for coincident and leading indicators are

usually based on two-steps procedures logically well distinguished, even if they can be performed simultaneously in some cases. The first step is devoted to the classification of variables according to their leading properties and to variable selection in order to reduce the dimension of the variables space. In this step data analysis techniques such as static and dynamic factor analysis and principal components, discriminant analysis, etc can be used. This will lead to the identification of some latent variables (e.g. factors, principal components) to be used in the second step. Alternatively, approaches such as the general to specific and the clustering analysis techniques can be used to extract those variables with the highest explanation power. In the second step, a variety of methods from simple aggregation ones with fixed weights to regression based, bridge equations, VARs, linear and non linear time series models, etc can be used to estimate the coincident and leading indicators.

Coincident and leading indicators can supply a wider range of information with respect to the methods presented in previous sections. The previous methods were mainly oriented to the estimation of period on period or year on year growth rates and, in a more limited case, of levels of the target variable. Coincident and leading indicators can be targeted, thanks to the use of particular transformations applied to the components series, to the delivery of information covering the simple growth of the reference variable (period on period and year on year growth rates), its trend cycle or growth cycle features, as well as the occurrence of turning points.