Predictions vs. Preliminary Sample Estimates:
the Case of Eurozone Quarterly GDP

Enrico D’Elia
Working Papers

The working paper series promotes the dissemination of economic research produced in the Department of the Treasury (DT) of the Italian Ministry of Economy and Finance (MEF) or presented by external economists on the occasion of seminars organized by MEF on topics of institutional interest to the DT, with the aim of stimulating comments and suggestions.

The views expressed in the working papers are those of the authors and do not necessarily reflect those of the MEF and the DT.

© Copyright:

2014, Enrico D’Elia

The document can be downloaded from the Website www.dt.tesoro.it and freely used, providing that its source and author(s) are quoted.

Editorial Board: Lorenzo Codogno, Mauro Marè, Libero Monteforte, Francesco Nucci, Franco Peracchi

Organizational coordination: Marina Sabatini
Predictions vs. Preliminary Sample Estimates: the Case of Eurozone Quarterly GDP
Enrico D’Elia (*)

ABSTRACT

Economic agents are aware to incur in a loss basing their decisions on their own extrapolations instead of sound statistical data, but the loss could be smaller than the one related to waiting for the dissemination of final data. A broad guidance in deciding when statistical offices should release preliminary and final estimates of the key statistics may come from comparing the loss attached to users’ predictions to the loss associated to possible preliminary estimates from incomplete samples provides. Also the cost of delaying decisions for many economic agents may support the dissemination of very early estimates of economic indicators even if their accuracy is not fully satisfactory from a strict statistical viewpoint. Analysing the vintages of releases of quarterly Euro area GDP supports the view that even very inefficient predictions may beat some official preliminary releases of GDP, suggesting that the current calendar of data dissemination deserves some adjustment. In particular, actual “flash” estimates could be anticipated, while some later intermediate releases are likely less informative for the users.

JEL CLASSIFICATION: C44, C49, C82, C83.
KEYWORDS: Accuracy, Data Dissemination, Eurozone GDP, Forecast, Preliminary Estimates, Timeliness.

(*) Ministero dell’Economia e delle Finanze, Dipartimento del Tesoro e ISTAT. E-mail: enrico.delia@tesoro.it
## CONTENTS

1  INTRODUCTION ................................................................................................................................. 3

2  THE ACCURACY OF PRELIMINARY SAMPLE ESTIMATES AND FORECASTS ............................. 6

3  THE COST OF DELAYING DECISIONS .............................................................................................. 13

4  ANALYSING THE RELEASES OF QUARTERLY GDP ESTIMATES FOR THE EUROZONE .......... 15

5  CONCLUDING REMARKS .................................................................................................................. 24

REFERENCES ............................................................................................................................................... 26
1 INTRODUCTION

The trade-off between accuracy and timeliness of statistical data is a key issue for statistical offices. It has been analysed mainly with reference to estimates of GDP and other “Principal European Economic Indicators” identified by the Economic and Financial Committee of the European Commission, aimed at detecting the turning points of the business cycle earlier. An international conference organised by the UNSTATS (2009) discussed in depth the same topic and the OECD analysed the quality of statistical information within the “Short-Term Economic Statistics Timeliness Framework”. Notably, Altavilla and Ciccarelli (2007) and the European Central Bank (2009) pointed out that the flash estimates of European GDP do not differ significantly from the official first releases published later, so that early estimates are probably more helpful for the decision makers than the corresponding final releases. Also economic agents form their informed predictions on the relevant variables while waiting for official data releases. The main aim of this paper is to show how the “competition” between the accuracy of users’ judgements and the accuracy of early official estimates may provide some guidelines for improving the data dissemination policy of statistical offices, particularly for the quarterly estimates of GDP in the Eurozone.

Early estimates of economic indicators are welcomed by decision makers who are not in the position to wait for the dissemination of the final results of the pertinent statistical surveys before choosing from alternative strategies. In particular, the timing is important in most decision processes concerning investment, consumption, price setting, etc. Thus, users of statistical data often have to resort to model based predictions on the final outcome of some statistical surveys on past and current facts, often referred to as “nowcasts”, to be distinguished from genuine forecasts about the future. In other words, predictions and preliminary results from surveys can be regarded by decision makers as imperfect substitutes. This fact doubtless offers a novel viewpoint on the trade-off between timeliness and accuracy of statistics, providing some suggestions about the strategy for disseminating statistical data. Particularly, the implicit competition between nowcasts and early estimates should be taken into account together with the usual assessments on production costs, technical capability, transparency, credibility and legal obligations of statistical offices.

1 The author gratefully acknowledges the valuable suggestions and criticisms which come from the referees of this journal, Alberto Zuliani and the participants of a series of seminars. Of course, errors and omissions are the responsibility of the author. A revised version of this paper is forthcoming on the Journal of Official Statistics.

2 Related documents are available at www.oecd.org/document/40/0,3343,en_2649_34257_30460520_1_1_1_1,00.html.
Data users are perfectly aware that the final results of statistical surveys are more accurate than forecasts, nowcasts and early estimates. In principle, profit expected from decisions based on very precise final statistical data is higher than profit deriving from choices founded on predictions and first releases of pertinent statistics. Nevertheless, waiting for the final results of statistical surveys before deciding, is costly as well, as profitable actions are postponed and economic resources are left unused, resulting in further costs. In addition, users know that both the accuracy of their predictions and of preliminary estimates usually improve over time, at least in “normal” conditions, when no major shocks hit the economy and the data collection process. Indeed, at the beginning of data collection, say at time $t$, users’ predictions are expectedly superior to any pure sample estimate, since the former embody public and private information, while the variance of pure survey estimates based on very few observations is virtually infinite, unless the statistical offices adopt an explicit Bayesian approach and reliable priors, which is infrequent in official statistics.

For their part, statistical offices acknowledge that earlier estimates meet the needs of most users, and are generally technically capable of producing excellent nowcasts, also by exploiting experts’ judgements and confidential sources of information. In principle, the statistical offices would be able to release a mass of preliminary data as well, even though they are aware it could cost more.. Nevertheless, official statisticians recognise that data revised too frequently and too much would confuse users and possibly damage institution credibility. In addition, publishing provisional data, possibly not included in the release calendars agreed to at international level, would raise uncertainty and search costs for users and introduces unduly informational asymmetries in international statistics which can ultimately impair users. Thus the current data release calendar hopefully finds the middle ground among many different requirements and constraints, and the specific viewpoint presented here should be correctly considered only as an additional one.

Let us assume that the accuracy of statistical estimates improves as data collection proceeds over time, achieving on average the accuracy of users’ forecasts only at time $t+h_0$, while information available to the users does not improve significantly. It follows that typical users would not exploit and appreciate figures possibly released before $t+h_0$, because these figures are considered less accurate than their own nowcasts. The threshold $h_0$ depends crucially on subjective users’ conditions, and mainly on their information and technical capability. Many users may wait intentionally for “official” data as long as preliminary estimates are expected to improve very fast, while the loss of making decisions based on inaccurate forecasts could be large. Also users’ extrapolations hardly beat preliminary survey results when a major shock hits the economy, making model based predictions inevitably less accurate. Nevertheless, the threshold $h_0$ is hardly null, and
may be quite large if the accuracy of early estimates does not sufficiently increase over time or even decreases in some cases.

Eurozone GDP estimates, analysed in the next sections, come from a complex procedure that exploits both pure sample information and model based estimators. As a consequence, comparing official GDP preliminary estimates and users’ nowcasts should provide strong evidence in favour of the dominance of official preliminary estimates, supporting the current dissemination policy of Eurostat, since the efficiency of the data elaboration process most likely reduces $h_0$ significantly. Nevertheless, the empirical evidence presented in Section 4 seems to show that even very inefficient predictions may do better than some preliminary estimates of GDP, suggesting that there is scope for improving the calendar of data releases even if representative data users are not very sophisticated. However, this result may be influenced by the particular period of time analysed (2002-2012) and by the small sample of fully comparable data available. Of course, a more comprehensive analysis of costs and benefits of changing the present calendar is also needed. In addition, conclusions depend crucially on the assumed ability of the representative users to form good forecasts and to exploit available information.

The next section exploits some well-known properties of preliminary estimates from incomplete samples to derive an ideal calendar for disseminating preliminary estimates exactly when their accuracy beats the errors size of model based predictions. The main conclusions are derived under the ideal condition that no large shock perturbs the economy and the accuracy of official estimates improves over time. The consequences of departing from this simplified framework are discussed briefly as well. The third section introduces the cost of delaying decisions while waiting for better official estimates. This issue, if taken into consideration, should encourage statistical offices to anticipate the release of data, but also clarifies that the dissemination calendar should also adapt to the characteristics of some “representative” user of statistical data, endowed with given capability and needs. Thus, acknowledging that statistical offices must serve different users, including legislators and governmental agencies, is crucial. The fourth section analyses the different vintages of quarterly GDP estimates in the Eurozone, regularly released by Eurostat, and recommends some adjustments to the current dissemination policy, even under the simplified hypothesis that users form very naïve predictions on GDP and do not incur costs for delaying their decisions. In particular, the suitability of the three major data releases currently available (respectively 45, 65 and 100 days after the end of the reference quarter) is discussed. Some concluding remarks close the paper.
2 THE ACCURACY OF PRELIMINARY SAMPLE ESTIMATES AND FORECASTS

Let $x_{i,t}$ measure a quantitative characteristic of the $i$-th individual at the time $t$, whose unconditional mean is $m_t$. Assumedly, the “representative” economic agent has to base his or her decisions on $m_t$ by using only the incomplete information set $\Omega_{t+h}$ available at time $t+h$. Typically $\Omega_{t+h}$ includes the past releases of the time series of $m_t$ and other aggregate economic indicators related to $m_t$; private information generally unavailable to the statistical offices; “soft” statistics, also produced by private agencies; judgements of experts. Nevertheless, $\Omega_{t+h}$ excludes the observations on $x_{i,t}$ collected and processed by the statistical office until $t+h$.

Thus, at least two provisional estimates of $m_t$ are ideally available at the time $t+h$:

(a) $f_{t+h} = E(m_t|\Omega_{t+h})$, that is the subjective prediction produced by exploiting the information set $\Omega_{t+h}$;

(b) $s_{t+h}$, that is the preliminary estimate based on the first $M_{t+h}$ observations collected at time $t+h$ by the statistical office.

Within this simplified framework, the representative user has the advantage of exploiting prior beliefs and private information, but has no access to individual records collected by the statistical office. The latter is allowed to use sample observations, but no other potentially useful pieces of information on $m_t$. In principle, statistical offices could develop mixed estimates within an explicit Bayesian framework, also taking into account experts’ judgements and other relevant non-sample information. Although the Bayesian approach has many theoretical advantages, it is infrequently used to improve sample estimates directly, because statistical offices tend to avoid estimation procedures that risk appearing too subjective, in view of defending and strengthening their neutrality and independence, in compliance with the first principle of the European Statistics Code of Practice.\(^3\) Although Little (2012) points out the possible advantage of adopting an explicit Bayesian approach in official statistics and discusses an application to the US Census data, Bayesian methods are applied in official statistics mainly to treat non-responses (see Graham, Young and Penny, 2009), to reduce the disclosure risk in the dissemination of individual data (see Little, Liu and Raghunathan, 2004), to match the units of different surveys statistically (see D’Orazio, Di Zio and Scanu, 2006), but not to improve preliminary estimates directly.

The time series $\{x_{i,t}\}$ can be decomposed as follows

\[
x_{i,t} = f_{t+h} + v_{t+h} + e_{i,t}
\]  

where \( v_{t+h} = m_t - f_{t+h} \) is an innovation process, with \( E(v_{t+h}|\Omega_{t+h}) = 0 \) and \( E(v_{t+h}^2|\Omega_{t+h}) = \phi_h^2 \) not depending on \( t \), even though the unconditional average of \( v_{t+h} \), say \( E(v_{t+h}) \), is not necessarily null; \( e_{i,t} \) is an idiosyncratic factor with \( E(e_{i,t}) = 0 \) and \( E(e_{i,t}^2) = \sigma^2 \). Notably, the two assumptions on \( e_{i,t} \) are quite standard, while the hypotheses on \( v_{t+h} \) could be violated if some time-specific factor changes systematically the predictability of the relevant events. For instance, forecast accuracy of GDP likely changes at the turning points of the business cycle or when some structural change makes the economic activity more or less erratic. In the latter case the time invariance of \( \phi_h^2 \) does not hold, while the variance of \( e_{i,t} \) does not change necessarily.

Let individual observations be collected and processed by the statistical office randomly, regardless of whether they are gathered almost continuously over time or in large batches, as commonly occurs. In this case the subscript \( i \) in [1] may denote the collection order of data, without any loss of generality. Thus the preliminary pure sample estimates of \( m_t \) at time \( t+h \) is

\[
\hat{s}_{t+h} = \sum_{i=1}^{M_{t+h}} w_{i,t,h} x_{i,t} \tag{2}
\]

where the weights \( w_{i,t,h} \) are such that \( \sum_{i=1}^{M_{t+h}} w_{i,t,h} = 1 \) for each \( t \) and \( h \). Under the previous assumptions on \( e_{i,t} \) in [1] and on the random collection of data, the average \( E(\hat{s}_{t+h}) \) evaluated over every possible sample of size \( M_{t+h} \) equals \( m_t \). Furthermore, if the individuals’ deviations from the average are mutually independent, the usual assumption \( E(e_i e_j) = 0 \) for \( i \neq j \) applies, so that the standard deviation of \( \hat{s}_{t+h} \) is

\[
\sigma_h = \frac{\sigma}{\sqrt{M_{t+h}}} \tag{3}
\]

in the simplest case of equally weighted observations.

Within a Bayesian framework, the estimator \( \hat{s}_{t+h} \) and its variance should take into account the priors on \( m_t \), so that \( \hat{s}_{t+h} \) would be a weighted average of the sample mean [2] and the mean of the assumed probability distribution of \( m_t \). Also, if the data are drawn from a normal population and the prior distribution of \( m_t \) is normal as well, the posterior variance of \( \hat{s}_{t+h} \) is

\[
\sigma_h = \frac{\sigma}{\sqrt{M_0 + M_{t+h}}} \leq \frac{\sigma}{\sqrt{M_{t+h}}} \tag{4}
\]

where \( M_0 > 0 \) measures the confidence on the prior, that is the ratio between the variance of \( e_{i,t} \) and the variance of the probability distribution assumed for \( m_t \). The same result holds for the Theil – Goldberger mixed least square estimator of \( m_t \), regardless of the probability distribution of data and
priors. The parameter $M_0$ in [4] can be interpreted as the size of the virtual sample from which the prior distribution of $m_t$ has been estimated.

According to [3], $\sigma_h$ is virtually infinite before the survey begins, since no observation has yet been collected and $M_{t+h}$ is null. [4] also implies that $\sigma_h$ peaks up to its maximum when $M_{t+h}$ equals zero, and is almost certainly large, unless the confidence of the statistical office on its priors is implausibly strong. In any case, during the survey, $M_{t+h}$ is a non-decreasing function of $h$, for instance: $M_{t+h} = M(h)$ with $\frac{dM}{dh} \geq 0$, regardless of the reference period $t$. It follows from [3] and [4] that $\frac{d\sigma_h}{dh} \leq 0$ holds at least in “normal” conditions in which data collected when $h$ takes some special values, are systematically biased and volatile. Noticeably, this could be the case when most influential units are surveyed just at the beginning and the end of the data collection process respectively, for instance because they are able to fill the questionnaires only according to a special calendar (e.g.: just after balance sheets or periodic reports to the stakeholders have been published). In such unlucky cases, $\sigma_h$ may even increase with $h$ during some phases of the survey process. The case in which $\frac{d\sigma_h}{dh} \geq 0$ will be discussed only briefly, since it would be even more supportive of the advantage of nowcasts over official preliminary estimates.

The profit loss associated to the use of preliminary estimates, say $S(h)$, can be assumed to be a non-decreasing function of $\sigma_h$, say $S(h) = L(\sigma_h)$ with $\frac{dL}{d\sigma_h} \geq 0$ and $L(0) = 0$. The function $L(\sigma_h)$ depends largely on the subjective conditions of data users and on the specific decision to be based on statistical data. In particular, the inaccuracy on a variable could be almost negligible in some cases, and potentially harmful in others. For instance, estimating correctly the level and dynamics of GDP is very important in deciding investment, but not export strategies. Nevertheless, the formal properties of $L(\sigma_h)$ exploited in the following sections are not influenced by such subjective factors. Notably, $L(\sigma_h)$ is not necessarily a linear transformation of the standard deviation of errors $\sigma_h$, and particularly could be flat for a wide range of $\sigma_h$. It implies that $S(h)$ and $L(\sigma_h)$ are not necessarily quadratic functions of errors, as often assumed. The main limitation of the relationship $S(h) = L(\sigma_h)$ is that data users are assumed to be adverse to positive and negative estimation errors equally, contrarily to what Granger and Pesaran (2000) argued. For instance, if statistical data are used to design fiscal policies, the government is more likely to be worried about overestimating GDP growth, since less income entails larger budget unbalances, due to larger social expenditure and lower tax revenues. Also, most firms acting in a competitive market fear overestimating a potential
market much more than underestimating it, since the former error incurs larger investment and financial costs for the same actual turnover. On the contrary, in oligopolistic markets, plants could be oversized intentionally to prevent the entry of possible competitors, so that entrepreneurs would be more adverse to underestimating market size. In general, the symmetric relationship \( S(h) = L(\sigma_h) \) can be considered a feasible approximation of the true loss function of the representative agent only for small error size.

The main advantage of relating \( S(h) \) to \( \sigma_h \) is that it makes it possible to compare users’ predictions and preliminary sample estimates, disregarding the specific functional form of \( L(\sigma_h) \), that is the nature of decisions to be made by the representative user. As \( \sigma_h \) is not a continuous function of \( h \), also \( S(h) \) may share this discontinuity. For instance, if data are collected in batches, \( S(h) \) is very likely a piecewise continuous function, in all probability characterized by sudden drops after each batch of data has been processed. In any case, \( S(h) \) is suitable to be estimated empirically by statistical offices from the track of data collection, and can be approximated by users from the revisions of data, compared to some benchmark release, which can be considered the ultimate estimation, hopefully closest to the true value of the relevant variables. Furthermore, assuming that \( S(h) \) is a non-decreasing function of \( \sigma_h \) implies that \( \frac{dS(h)}{dh} \) shares the same sign of \( \frac{d\sigma_h}{dh} \), apart from possible discontinuities. For instance, Table 1 and Figure 1 in Section 4 provide some empirical evidence on the negative relationship between \( \sigma_h \) and the dissemination delay \( h \) of the preliminary estimates of quarterly GDP in the Eurozone released by Eurostat, compared to the official estimate released 400 days after each reference quarter. It is worth noticing that in the case examined here the condition \( \frac{d\sigma_h}{dh} \leq 0 \) holds even before the latest economic crisis, when the heterogeneity among the performances of firms was assumedly smaller. In addition, the accuracy of preliminary estimates of GDP seems to improve at decreasing rates as if the data elaboration process is much more efficient at the beginning of the statistical survey and each additional observation gives only a minor contribution to the accuracy of the sample estimates.

Since \( \Omega_{t+h} \supseteq \Omega_{t+h-1} \) by definition, it follows that \( \frac{d\phi_h}{dh} \leq 0 \), at least on average and in “normal” times, that is when news available at time \( t+h \) prevail on “noise”, as questioned by Blanchard et al. (2009). The function \( \phi_h \) can be also discontinuous, with sudden drops when some influential piece of information is usually available only when \( h \) takes some special values. The assumption \( \frac{d\phi_h}{dh} \leq 0 \), apart from some possible discontinuity points, relies crucially on the fact that the representative
user is able to keep, or hopefully to improve over time, its capacity to understand and exploit available information efficiently. Notably, full rationality of economic agents is not strictly required for \( \frac{d \phi_h}{dh} \leq 0 \). For instance, it is enough that they are “rationally inattentive” as argued by Sims(2003), that is, they intentionally disregard part of available information because collecting and elaborating it exceeds the profit expected from adjusting their decisions further. In any case, we will see that the hypothesis \( \frac{d \phi_h}{dh} \leq 0 \), although very likely and desirable, is not strictly necessary in designing an ideal calendar for data release.

Similarly \( \sigma_h \), also \( \phi_h \) can be measured empirically, for instance from direct surveys on users’ judgements, or assuming a reasonable mechanism for the formation of nowcasts, as done in Section 4. Given the relation between the expected profit loss and the accuracy of data used to make a decision, also \( F(h) = L(\phi_h) \) can be defined. Thus \( F(h) \), similarly to \( S(h) \), can be considered a non-decreasing transformation of errors’ size at time \( t+h \). This property allows us to compare \( \sigma_h \) and \( \phi_h \) instead of the subjective and unknown functions \( F(h) \) and \( S(h) \).

If \( v_{t+h} \) and \( e_{i,t} \) are not correlated, as assumed above in “normal” times, the decomposition [1] implies that

\[
E(\sigma^2_h) = E\left[ \frac{1}{M_{t+h}} \sum_{i=1}^{M_{t+h}} (x_{i,t} - f_{i,t+h}) \right] - \phi^2_h
\]

[5]

where the \( E(.) \) operator applies to the time series of the relevant variables. The expression in square brackets in [5] is larger than \( E(\sigma^2_h) \), since only the arithmetic average \( m_t \) minimizes the sum of squared discrepancies \( (x_{i,t} - f_{i,t+h}) \), thus \( \phi^2_h \) can be seen as the difference between the estimated variance among observations around the forecast \( f_{t+h} \), on one side, and the variance \( \sigma^2_h \) around the true average \( m_t \), on the other. Therefore \( \phi^2_h \) is most likely small compared to \( \sigma^2_h \), as long as \( f_{t+h} \) is a reasonable forecast of \( m_t \).

As noted above, rational agents are assumedly able to make forecasts even before data collection has begun, when \( \sigma^2_h \) is virtually infinite, so that \( \phi^2_h < \sigma^2_h \) for \( h \leq 0 \). As time goes on, predictions may improve, thanks to the availability of other relevant pieces of information, but probably at a slower pace compared to a survey. Otherwise, forecasts would do better than statistical surveys all the time and paradoxically the latter would have only a little value for the representative user.

Excluding the latter implausible case, the assumptions that \( \phi^2_h < \sigma^2_h \) for \( h \leq 0 \) and \( \frac{d \sigma_h}{dh} \leq \frac{d \phi_h}{dh} \).
subsequently (disregarding possible singular discontinuities) imply that \( \phi_h^2 = \sigma_h^2 \) at some point in time, say \( t+h_0 \). Noticeably, the condition \( \frac{d \sigma_h}{dh} \leq \frac{d \phi_h}{dh} \) ideally does not require that both \( \phi_h \) and \( \sigma_h \) are non-increasing function of \( h \). Furthermore \( h_0 \) could be very large, so that nowcasts could keep their advantage for a long while. What is really crucial is that the initial advantage of nowcasts (if any) tends to reduce as the dissemination delay increases.

If statistical offices actually release the sample estimates available before \( t+h_0 \), users are likely to continue basing their decisions on their own forecasts in order to minimise their expected profit loss, unless this new piece of information improves \( \Omega_{t+h_0} \) because users can embody in their newcasts even the very inaccurate preliminary data released before \( t+h_0 \). The expected loss associated to these “data adjusted” nowcasts determines a downward shift of the \( \phi_h \) function and a new intersection point between \( \phi_h \) and \( \sigma_h \), say at \( h = h_1 \). If the downward shift of \( \phi_h \) is substantial, economic agents would welcome even earlier provisional releases of data by the statistical office. In contrast, if intermediate data releases improve users predictions only to a lesser extent, users will continue to base their decision on their own past nowcasts even after the dissemination of official data in the meantime. As a consequence, comparing the functions \( \phi_h \) and \( \sigma_h \) after each data release may provide operative guidelines to refine the dissemination plans of statistical offices. In particular, the first data release could be anticipated if it causes a large downward shift of \( \phi_h \), even if the inaccuracy sample estimates is large. On the other hand, intermediate preliminary estimates that do not improve \( \phi_h \) enough should be avoided, since they are very likely costly for the statistical offices and less appreciated by the users. It is worth noticing that \( \phi_h \) and \( \sigma_h \) can be compared also if they present some discontinuities, thus the approach proposed here to design an ideal data release calendar seems quite general. For instance, Table 3 and Figure 2 in Section 4 show an example of the interplay between the publication of preliminary estimates and the elaboration of users’ nowcasts based on simple univariate time series extrapolations of quarterly GDP in the Eurozone. In particular, the dashed line in Figure 2 represents the accuracy of nowcasts adjusted after each data release that improves almost every time new data are published.

Indeed, the downward shift of the function \( \phi_h \) would be null if statistical offices provide the best nowcast by applying efficient model-based estimators to the collected data, so that users’ forecasts could hardly better the preliminary data published at time \( t+h \). The improvement in nowcasts can be seen as a special case of efficiently exploiting the data collected up to \( t+h \) by integrating missing data in the full sample by means of a model based estimator, as discussed in Sarndal (2005). In any case, users can only combine available forecasts, as suggested by Clemen (1989) and Yang and Zou
(2004), while the statistical offices possibly may combine the same forecasts and the provisional results of their surveys.

Provided that the survey ends at $t+H$, the relative performance of the two estimators $s_{t+h}$ and $f_{t+h}$ depends on the time schedule of the survey, which determines the coverage ratio $\frac{M_{t+h}}{M_{t+H}}$, on one hand, and the ratio $\gamma_h = \frac{\phi_h}{\sigma_H}$ of the mean square error of prediction to the variance of $m_t$ among observations at the end of the survey, on the other. The ratio $\gamma_h$ ranges from 0 to infinity: in particular, $\gamma_h$ is null if the time series $m_t$ is purely deterministic and tends to infinity if individuals are identical. For instance, the changes of the average age of a stationary population can be virtually predicted without any error even though the age differs among individuals very much. Contrarily, the yield of a homogeneous set of equities can hardly be predicted even if they have the same market price.

As assumed above cautiously, let forecast accuracy improve over time less than $\frac{\sigma_h}{\sigma_H}$, that is less than $\sqrt{\frac{M_{t+H}}{M_{t+h}}}$ according to [3]. Since rational agents prefer their own forecast to preliminary estimates of $m_t$ as long as $\sigma_h \geq \phi_h$, it follows that

$$\frac{M_{t+h}}{M_{t+H}} \geq \frac{\sigma_H}{\phi_h} \tag{6}$$

The inequality [6] has a number of interesting consequences. First of all, it implies that the subsample estimator is more efficient after some threshold $h$ only if $\gamma_0$ is not null, otherwise a rational agent would always be better off by making decisions based on his own nowcasts. Conversely, the preliminary results from incomplete samples are the best choice, at any time, only in the limiting case in which even one single observation provides better information than any forecast, so that $\sigma_h$ is null for whatever small dissemination delay $h$. Secondly, the threshold $\frac{M_{t+h}}{M_{t+H}}$ for making the publication of preliminary results valuable may be unexpectedly large even when the prediction accuracy is quite poor compared to final sample mean variance $\sigma_H$. For instance, if $\phi_0$ is as (implausibly) large as ten times $\sigma_H$, the minimum subsample for data publication would be larger than 10% of the complete sample.
3 THE COST OF DELAYING DECISIONS

Other than the cost of taking decisions on inaccurate data, often economic agents have to also consider the additional cost of delaying decisions, as argued by Granger and Machina (2006). This is the typical case when the “first mover” has some advantage over the followers. For example, if the potential market is given, the first firm entering the market may hopefully serve the most profitable segment of demand, while the followers have to supply only the others. Also purchasing and investment decisions are usually supposed to have an optimal timing, mainly related to economic fluctuations. Winston (2008) provided a comprehensive survey of economic models in which decision timing is a major factor.

In some special cases, taking into account the cost of delaying decisions may imply that users incur smaller overall losses if they based their decisions on timely but very inaccurate nowcasts instead of delayed preliminary and final official estimates of the relevant variables. In fact, the loss of delaying decisions may be so fast growing over time that agents cannot afford to wait for more accurate but late survey results.

The cost of delaying decisions, waiting for more accurate information, is presumably a function of time passed from the reference period of relevant information, say $D(h)$. The function $D(h)$ achieves its minimum at $h=0$, when assumedly $D(0) = 0$ without any loss of generality, and the cost of delaying decisions very likely does not decrease with $h$, that is $\frac{dD}{dh} \geq 0$.

Let $F(h)$, $S(h)$ and $D(h)$ be continuous functions of $h$ only to make the problem more tractable analytically, and suppose that $F(h)$ and $S(h)$ cross for the first time at the delay $h_0$, as assumed in Section 2, rational agents exploiting only predictions incur the minimum overall loss $L_f$ at $h_f$, that is approximately

$$L_f = (F(h_c) + D(h_0)) + (f' + d')(h_f - h_0) + (f'' + d'')(h_f - h_0)^2$$

while those basing their decisions on the preliminary results of surveys face the minimum loss, say $L_s$, $h_s$ periods after the reference time, that is about

$$L_s = (S(h_c) + D(h_0)) + (s' + d')(h_s - h_0) + (s'' + d'')(h_s - h_0)^2$$

where $x' = \left. \frac{dX}{dh} \right|_{h=h_0}$ and $x'' = \left. \frac{d^2X}{dh^2} \right|_{h=h_0}$.

According to [7] and [8], the losses $L_f$ and $L_s$ achieve their minima when

$$h_f = h_0 - \frac{1}{2} \frac{f' + d'}{f'' + d''}$$

and

```
13
```
\[ h_s = h_0 - \frac{1}{2} \frac{s' + d'}{s'' + d''} \]  \[ \text{that is when} \]
\[ L_f = (F(h_0) + D(h_0)) - \frac{1}{4} \frac{(f' + d')^2}{f'' + d''} \]  \[ \text{and} \]
\[ L_s = (S(h_0) + D(h_0)) - \frac{1}{4} \frac{(s' + d')^2}{s'' + d''} \]

In principle, according to [11] and [12] the minimum loss could be achieved either basing decisions on forecasts or on sound statistical data, depending on the shape of the functions \( S(h), F(h) \) and \( D(h) \). Indeed, since \( F(h_0) = S(h_0) \) by definition, the condition for \( L_f \leq L_s \), together with [11] and [12], imply that
\[
\frac{(s'+d')^2}{s''+d''} \geq \frac{(f'+d')^2}{f''+d''}
\]
Assuming that users’ nowcasts improve over time only very slowly, and much slower than the results of surveys, \( f' \approx 0 \) and \( s'' \geq f'' \). Thus [13] also means that
\[
s' \geq -2d'
\]
As a consequence, even under assumptions very unfavourable to the accuracy of users’ forecasts, the representative agents would be better off basing their decisions on some predictions, instead of waiting for the preliminary and the final results of relevant surveys, if the marginal improvement of survey accuracy, that is \(-s'\), does not exceed twice the loss attached to postponing decisions one unit of time more, that is \(d'\). The condition [14] fully conforms the intuition, according to which agents prefer basing their decisions on predictions as the cost of delay increases very fast, and the expected error size of surveys does not fall too fast over time.

Of course, the condition [14] does not necessarily hold, so that, in general, agents hopefully find it less costly to wait for the dissemination of the final or even preliminary results of statistical surveys. It is worth noticing that the result [14] does not take into consideration the possibility that disseminating preliminary survey results might dramatically increase the accuracy of forecast. Otherwise, it could happen that the minimum loss associated to predictions is always lower than that deriving from making decisions based only on survey results, since, in this case, the curve \( F(h) + D(h) \) lies below \( S(h) + D(h) \) by definition.

Unfortunately, \( D(h) \) cannot be related to \( \sigma_h \) or \( \phi_h \), contrarily to \( S(h) \) and \( F(h) \), thus the condition [14] strictly depends on the specific decision problem faced by economic agents. Thus this factor is not considered in the next section. Nevertheless, [14] implies that users may appreciate preliminary
estimates released much earlier than $h_0$, when the accuracy of sample estimates crosses the accuracy of users’ predictions.

4 ANALYSING THE RELEASES OF QUARTERLY GDP ESTIMATES FOR THE EUROZONE

In the European Union quarterly national accounts are released according to a “minimal” calendar established by the EC Regulation N° 1392/2007. However the statistical offices of the member states and Eurostat tend to provide data even more timely than prescribed by this Regulation. At the moment, three main releases are published for each quarter:

1. the first release 45 days after the end of the reference quarter is named “flash estimate” and consists of GDP growth estimates for the latest quarter only. No component of GDP is published at this stage;
2. the “second release” about 65 days after the end of the reference quarter, including a basic breakdown GDP. A more complete set of data, including an estimate of domestic employment, follows about ten days after the “second release”;
3. the “third release” is scheduled at around 100 days after the end of the quarter. It provides more detailed breakdowns for the latest quarter.

Quarterly data are open to backwards revision at each release, and data on the previous three years are usually subject to major revisions when annual data are released by March for the “excessive deficit notification” prescribed by the European rules. Furthermore, seasonal adjustment procedures may bring some minor revisions of quarterly data even older than three years. As a consequence, many different “vintages” of GDP estimates are available for each quarter: combining the sequence of the three releases listed above, the GDP estimate for a given quarter is possibly subject to 11 revisions during the subsequent 12 months.

It is worth noting that the national account estimates derive from a very sophisticated process that exploits both the results of pure preliminary sample estimates on a large number of statistical indicators, and a range of model based procedures aimed at integrating missing data and treating possible outliers. Thus, quarterly GDP vintages almost certainly improve their accuracy over time much faster than a sequence of pure non-Bayesian estimates from incomplete samples like that considered in Section 2 for illustrative purposes. Thanks to the mass of non-sampling information embodied in each release of data, the $\sigma_h$ function associated to the GDP vintages decreases

---

expectedly faster than a pure sample estimate, so that the comparison between \( \sigma_h \) and \( \phi_h \) is very unfavourable to users’ predictions at any time. The comparison is even less favourable to users’ predictions if the growth rates in the same quarter of the previous year are considered, since this transformation of original time series tends to reduce two major sources of revision, i.e.: best information on the level of GDP, mainly related to back-revision of annual data, and the changes of data induced by running seasonal adjustment procedures on longer time series.

The different “vintages” of year-on-year growth rates of volume GDP, seasonally and working day adjusted, for the 12 countries of the Eurozone are collected and published regularly on the website of Eurostat starting from the rate of 2003Q1. Older data are considered much less comparable over time and across the member states. As of the end of 2012, the last available data on GDP revisions refer to 2011Q4 because final releases of later data are unavailable.

The so called “triangle of revisions” published by Eurostat shows that the largest revisions of GDP estimates occur within 6-9 months after the reference quarter, but in principle GDP can be revised many times for about 3 years after the reference quarter, following the regular revisions of annual data. In addition, seasonal adjustment procedures may induce further minor changes of data even after 3 - 4 years. However, no economic agent is probably in the position to wait for such a long period of time before making a decision, thus here the benchmark for evaluating the accuracy of preliminary estimates has been set arbitrarily to 400 days after the reference quarter (that is after about 13 revisions), also to save degrees of freedom to carry out further statistical analysis. A comparison of real time data with their 3rd year benchmark (corresponding to the latest release admitted for the excessive deficit notification) will be discussed briefly below.

The revisions of GDP represent a challenging case study for simulating the interplay between users’ nowcasts and official data releases sketched in Section 2. Since this paper aims at testing the possible advantages of users’ estimates over current official estimates, a number of assumptions unfavourable to users’ nowcasts have been adopted throughout the simulation exercise. In particular, the revision of annualized growth rates of GDP are considered, and users’ estimations are simulated by using intentionally simple and inefficient procedures that exclude any piece of information other than the time series of GDP vintages.

Table 1 reports some statistics on the accuracy of preliminary estimates of GDP in the Eurozone evaluated vis à vis the 400 days benchmark estimated on the sample 2003Q1 to 2011Q4 and on the pre-crisis subsample ranging from 2003Q1 to 2008Q2. The latter is part of a period sometimes called the Great Moderation, because business cycle fluctuations, and average growth rate of GDP,

---

5 The revision triangle can be downloaded from http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/methodology/quarterly_accounts in excel format
were very weak. Thus both preliminary estimates and nowcasts were exposed only to minor unpredictable shocks. On the contrary, the post 2008 sample includes the largest economic crisis after World War II, and has provided many surprises to forecasters and statisticians.

Table 1 – The accuracy of preliminary estimates of GDP

<table>
<thead>
<tr>
<th>Dissemination delay</th>
<th>Average error</th>
<th>RMSE</th>
<th>5th centile</th>
<th>95th centile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full sample</td>
<td>Until 2008Q2</td>
<td>Full sample</td>
<td>Until 2008Q2</td>
</tr>
<tr>
<td>45 days</td>
<td>-0.024</td>
<td>-0.052</td>
<td>0.209</td>
<td>0.179</td>
</tr>
<tr>
<td>65 days</td>
<td>-0.031</td>
<td>-0.058</td>
<td>0.159</td>
<td>0.132</td>
</tr>
<tr>
<td>100 days</td>
<td>-0.022</td>
<td>-0.041</td>
<td>0.137</td>
<td>0.123</td>
</tr>
<tr>
<td>101 - 200 days</td>
<td>-0.022</td>
<td>-0.036</td>
<td>0.098</td>
<td>0.089</td>
</tr>
<tr>
<td>201 - 250 days</td>
<td>-0.018</td>
<td>-0.026</td>
<td>0.073</td>
<td>0.071</td>
</tr>
<tr>
<td>251 - 350 days</td>
<td>-0.011</td>
<td>-0.014</td>
<td>0.040</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Although the figures reported in Table 1 should be considered cautiously because only 24 degrees of freedom are available for the computation of statistics, some evidence is reasonably clear. First of all, preliminary estimates show a weak downward bias in both periods, although not significant from the pure statistical point of view, possibly because the statistical offices are usually more concerned with overestimating GDP growth rates rather than with revising upward the data during the following years. Notably, this evidence was even stronger before the last economic crisis, supporting the view that the accuracy of official estimates has not been influenced too much by the large adverse shocks that hit the economy after 2008Q2. In any case, the negative bias tends to vanish as the delay of preliminary estimates increases from 45 days to 250 days and over.

Also Table 1 shows that the root mean square error (RMSE) of preliminary estimates decreases quite fast, as conjectured in Section 2: in the full sample it falls from 0.21 percentage points for the flash estimates to 0.04 percentage points for the oldest vintage considered here; before the crisis the RMSE ranged from 0.18 to 0.04, that is not much lower than the same statistic calculated for the full sample of data. This evidence supports the hypothesis that the preliminary estimates of GDP are very robust to large shocks. Furthermore, in 9 cases out of 10, between 2003 and 2011, the revisions range from -0.34 to 0.27 percentage points for the flash estimates, and only from -0.07 to 0.06 percentage points for the 250-350 days releases, and similar results also come from the analysis of the pre-crisis period.

The same evidence is confirmed by the non-parametric estimate of the function $\sigma_h$ reported in Figure 1, even taking into account again that few degrees of freedom are available particularly for the estimation on longer dissemination delays. The local second degree polynomial estimator described by Fan (1992) has been adopted. For each vintage, the interpolation is based on a series of
weighted least square estimators in which the observations close to the reference vintage are weighted by a “kernel function”. The “bandwidth” of the weighted observations has been determined according to the formula proposed by Fan and Gijbels (1996). The main drawback of this methodology is that it assumes the continuity of interpolated functions that could be unrealistic, as argued in Section 2.

Also according to non-parametric analysis, the RMSE of revisions decreases with the dissemination delay both during the pre-crisis period and the full period 2003 – 2011, even though the decline is faster during the first 60 - 90 days and slower afterwards, supporting the view that the statistical offices are able to exploit the most informative data by the beginning of the estimation process. The virtual RMSE of preliminary data released just at the end of the reference quarter would be 0.33 percentage points, that is almost 50% larger than the actual RMSE of flash estimates. Nevertheless, this value is likely underestimated, since it definitively comes from a pure backwards extrapolation of the observed rate of changes of $\sigma_h$ between 45 and 65 days after the end of the reference quarter, and is not consistent with [3] and [4]. In any case, the RMSE of preliminary estimates apparently halves within about 80 days, regardless of the period considered, and divides by four within about 180 days.

**Figure 1 – The accuracy of quarterly GDP preliminary estimates**
Most results found comparing GDP revisions to their 400 day benchmarks are confirmed by considering the 3 year benchmarks instead, although the degrees of freedom for estimating RMSE and other statistics drop dramatically. In particular, $\sigma_h$ is still decreasing as the dissemination delay increases although the RMSE of flash estimates picks up to 0.371, that is about 80% more than the RMSE computed versus the 400 days benchmark. The latter benchmark is still subject to revisions whose average size is 0.252 percentage points during the next 600 days. The non-parametric estimation shows that the $\sigma_h$ curve is almost parallel to the one computed for the 400 day benchmark, beyond the second release of data. Detailed descriptive and non-parametric statistics for the 3 year benchmark are not reported here for sake of brevity, and are available from the author.

In order to compare the pure statistical estimates to users’ predictions and nowcasts, a forecasting model has been assumed. To make the exercise more challenging, the forecasts made before the end of each reference quarter and the adjusted nowcasts based on the following preliminary estimates are simulated by using intentionally very simple time series models, estimated inefficiently by means of ordinary least squares on real data available at the moment of each simulation. This procedure intends to mimic the actual behaviour of an unsophisticated user who exploits only official information readily available on GDP and disregards any other evidence, such as timely short term statistics, “soft data” on business and household confidence, possible private information, etc. Thus, in principle, the experiment is strongly biased toward the superiority of official estimates and, in principle, should support the actual data dissemination policy adopted by Eurostat, since subsequent official estimates potentially embody more information than that used by the imaginary naïve user considered in our simulation.

Some insight on actual accuracy of forecasts and nowcasts on GDP made by more sophisticated users is provided by Barhoumi et al. (2008), Diron (2008), Angelini et al. (2011) and Frale et al. (2011), who developed very short term forecasts and nowcasts of Eurozone GDP, and by Pain and Sédillot (2005), who applied similar methods to other OECD countries. Joint nowcasts and short term forecasts of inflation and GDP were proposed by Giannone, Reichlin and Small (2008). In those papers, the RMSE of nowcasts based on real time information likely available to data users ranges from 0.6 and 0.2 percentage points, with a gain of using additional information peaking up to even 40% of naïve predictions. Also Jansen, Jin and de Winter (2012) estimated that consensus forecast collected by ECB among experts are slightly more accurate, with a further cut of $\phi_h$ by 10% compared to the best statistical models.

In this simulation experiment, the forecast $Y^*_{t_i}$ on the yearly growth rate of GDP made before the end of the reference quarter $t_i$ derives from the simple AR model

$$Y^*_{t_i} = c_{t, v} + a_{t, v} Y_{t-i, v} + u_{t, v}$$

[15]
where $c_{t,v}$ and $a_{t,v}$ are parameters estimated by using only the latest data available at time $t-1$, not including $Y_t$ thus far; $Y_{t-1,v}$ is the latest release of the GDP growth rate at time $t-1$; $u_{t,v}$ is a random disturbance, likely autocorrelated, being a forecasting error, and possibly heteroscedastic. Even though the assumed characteristics of $u_{t,v}$ would require appropriate methods for estimating [15] efficiently, our imaginary user is supposed to use only ordinary least squares. In any case, scarce degrees of freedom available for the simulation (on occasion less than 10) would make it unfeasible to use other proper estimation methods. This practice makes forecasts even worse than those possibly produced by the model [15] itself. In order to allow some degree of freedom to the estimates, the results of the first ten regressions have been discarded and the first forecasting period has been set to 2006Q1. The first column of Table 2 reports the main results of the regression run to predict GDP at the end of the simulation period. It is apparent that the naïve model [15] fits the data quite well, even though the RMSE is as large as the average yearly growth of GDP during the last decade, mainly due to very few large outliers. Furthermore, forecasts tend to revert to the average after each deviation, as the estimate of the parameter $a_t$ is significantly below 1, thus, in principle, the model is incapable of predicting sudden turning points of GDP growth rate correctly. The coefficient of the dummy variable is not significant, at least for the last estimation period. The results of the regressions run, to produce the forecasts and nowcasts for each point in time, are not reported here, and are available on request.

Table 2 – The main results of regressions used to simulate users’ nowcasts and forecasts

(\textit{the statistics refer only to the longest sample available for each model})

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Forecast one quarter ahead</th>
<th>Dissemination delay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>45 days</td>
</tr>
<tr>
<td>$Y_{t,v-1}$</td>
<td></td>
<td>1.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>$Y_{t-1,v-1}$</td>
<td></td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.155)</td>
</tr>
<tr>
<td>$Y_{t-1,v}$</td>
<td>0.895</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>Dummy variable</td>
<td></td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.074)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.133</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.199)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.795</td>
<td>0.999</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.036</td>
<td>0.072</td>
</tr>
</tbody>
</table>

\textit{Standard error of estimates are under the coefficients in parentheses.}
When a new release of GDP figures, say $Y_{t,v}$, is published, users can improve their nowcast of the GDP growth taking into account also the previous revisions and the past dynamics of GDP. In this exercise, this “adjusted” estimate, say $Y^*_{t,v}$, has been simulated by using the model

$$
Y^*_{t,v} = c_{t,v} + a_{t,0,v-1} Y_{t,v-1} + a_{t,1,v-1} Y_{t-1,v-1} + a_{t,1,v} Y_{t-1,v} + d_{t,v} D_t + v_{t,v}
$$

[16]

where the parameters are estimated on the sample of data actually available when the $(v-1)$th vintage of data is released; $D_t$ is a dummy variable that is 1 at time $t-1$ and zero elsewhere, which serves to “sterilize” the forecast from the interpolation error made at time $t-1$. The rationale for [16] is that the revisions of GDP are hardly purely random and serially uncorrelated, so that there is room for improving the accuracy of the official estimates by also taking into account the typical time series structure of revisions. Similar evidence is reported by Fixler and Grimm (2006) also for the US GDP, and by Frale and Raponi (2012) in the case of Italy.

The main results of estimating the models [15] and [16] for the largest available samples are reported in Table 2. It is apparent that every model fits the data quite well, but there is strong evidence that the models are over-parameterized. In fact, regression results show that only the coefficients of $Y_{t,v-1}$ are statistically significant. In addition it is higher than 1 at any reasonable confidence level confirming the tendency of statistical offices to revise GDP growth upwards at each release of data. Apparently the underestimation is fairly substantial, ranging from 6% for the flash estimates to 3% for the 100 days releases. Contrarily, other regressors are not statistically significant: this result was expected for the dummy variable that serves only to sterilize the effects of possible outliers in the most recent estimation period, while it is unexpected for $Y_{t-1,v-1}$ and $Y_{t-1,v}$.

Indeed, the few available degrees of freedom of estimates and the strong collinearity between the regressors may mask the true influence of those variables. In fact, excluding them from the regressions significantly worsens the accuracy of adjusted nowcasts.

### Table 3 – The accuracy of nowcasts and adjusted preliminary estimates of GDP

<table>
<thead>
<tr>
<th></th>
<th>Average error</th>
<th>RMSE</th>
<th>5th centile</th>
<th>Median</th>
<th>95th centile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pure forecast one quarter ahead</strong> (a)</td>
<td>-0.012</td>
<td>0.262</td>
<td>-0.449</td>
<td>-0.015</td>
<td>0.542</td>
</tr>
<tr>
<td><strong>Adjusted preliminary estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 days</td>
<td>-0.013</td>
<td>0.147</td>
<td>-0.239</td>
<td>-0.020</td>
<td>0.189</td>
</tr>
<tr>
<td>65 days</td>
<td>-0.004</td>
<td>0.148</td>
<td>-0.165</td>
<td>-0.049</td>
<td>0.182</td>
</tr>
<tr>
<td>100 days</td>
<td>0.004</td>
<td>0.130</td>
<td>-0.137</td>
<td>-0.025</td>
<td>0.205</td>
</tr>
</tbody>
</table>

The statistics are computed on the sample 2006Q1 – 2011Q4 to have at least a ten observations for running each regressions.

(a) Excluding only the large forecast error on 2009Q2 (-2.931). Considering the full sample, the average error is -0.234 and the RMSE is 1.444.
The overall performance of one-step ahead forecasts and adjusted nowcasts are summarised in Table 3. The most interesting result is that, excluding a single large forecast error in 2009Q2 (about 3 percentage points below the true value), the predictions made before the end of the reference quarter are unexpectedly accurate, although they are intentionally naïve extrapolation-based. In fact, simulated users’ forecasts are less downward biased than most preliminary estimates and exhibit a RMSE that is roughly comparable to flash estimates. This evidence deserves further attention, since the simulation period comprises the data on the last global crisis, where large unexpected shocks hit the European economy and a “double dip”, including 3 turning points, occurred. However, model [15] also produced a number of large positive and negative errors, compared to the preliminary official estimates, as confirmed by the value of the 5\textsuperscript{th} and 95\textsuperscript{th} centiles of the distribution of errors that almost doubled the corresponding statistics computed for the flash estimates. As a result, comparing the accuracy of the simulated forecasts to non-parametric interpolation of \( \sigma_h \) it emerges that even naïve users’ predictions would compete against preliminary estimates of GDP possibly released about 30 days after the end of the reference quarter. This is really surprising, even taking into account that the out-of-sample interpolation of \( \sigma_h \), likely underestimates the accuracy of estimate when \( h \) is below the first dissemination delay actually observed. The horizontal piece of the dashed line in Figure 2 shows how early the simulated \( \phi_h \) function crosses the \( \sigma_h \) function for the first time.

**Figure 2 – Comparing the accuracy of nowcasts and preliminary estimates**

![Graph showing the comparison of preliminary estimates and nowcasts](image-url)
As argued in Section 2, this is a situation in which very timely official releases of GDP data, for instance just several weeks after the end of the reference quarter, would not be so “competitive” from the point of view of a representative economic agent. Nevertheless, if Eurostat decided to release such data, users might exploit this new piece of information to elaborate even better nowcasts, hopefully surpassing their previous projections one step ahead.

In fact, the second row of Table 3 suggests that when flash estimates are published, users are able to very much improve the accuracy of their adjusted nowcasts. The RMSE of nowcasts based on flash estimates is placed just amid the RMSEs of the official estimates of GDP released respectively 65 days and 100 days after the reference period. More precisely, Figure 2 suggests that after the flash estimates the \( \phi_h \) function shifts downwards substantially and intersects \( \sigma_h \) function when the dissemination delay is 95 days. Contrarily, when the 65 day official estimates are released, users’ adjusted nowcasts do not improve so much, as is apparent from the third row of Table 3. Therefore, the “second” release of GDP data has very likely only a minor impact on users’ decisions based on the dynamics of output in the Eurozone. However, the 65 days release of data includes a breakdown of data that expectedly improve the information set available to economic agents; thus the second release of GDP is welcomed by users focusing on sectorial dynamics rather than on the overall economic performance of the Eurozone.

The publication of the third release of quarterly GDP, 100 days after the end of the reference quarter, seems to increase the accuracy of users’ nowcasts further, as Figure 2 and the last row of Table 3 make evident. Nevertheless, the improvement is relatively too small to change users’ decisions significantly, so that they could became “rationally inattentive” as argued by Sims (2003), even if the accuracy of later official data releases of GDP is expectedly increasing. In any case, evidence for longer dissemination delays could be influenced by the scarce degrees of freedom available for estimation.

To sum up, the approach proposed in Section 2 and the empirical evidence presented in this section suggest producing a very early estimate of GDP as soon as possible before the first official release, possibly after few weeks, followed by a second release only 3-4 months later. Noticeably, the thresholds above were determined assuming that the typical user of data does not make use of very sophisticated forecasting methods and large information sets, and that after each nowcast, made when official data are disseminated, the nowcast does not improve further. Otherwise, the horizontal pieces of the dashed line in Figure 2 would be downward sloped, so that \( \phi_h \) would cross the curve of the accuracy of official releases later than 30 or 100 days after the end of the reference quarter. In addition, the cost of waiting, considered in Section 3 could prompt Eurostat to
disseminate even more timely data to better meet the needs of users that are not in the position to wait too much time before making their decisions.

A parallel simulation exercise, carried out on data revisions versus the 3 year benchmark, provided very similar results, although taking into account the drop in the degrees of freedom available to simulate users’ nowcasts. Indeed, this outcome was almost fully expected, since the revisions made during the first 400 days are most likely uncorrelated to those occurring in the following two years, which are related mainly to the availability of very detailed structural information available only after years, and are likely less correlated to the short term indicators mostly used to compute earlier estimates of GDP. Thus the size of revisions over the two benchmarks differ almost for a constant term, roughly explained by the difference between the 400 day estimate and the “definitive” 1000 day data release, as remarked on above. Given that users’ nowcasts cannot depend on data available only in the future, their accuracy versus the definitive data worsens only by a constant term as well, so that the relative comparison versus the accuracy of official data is almost unchanged. Full details of this experiment are available from the author.

5 CONCLUDING REMARKS

Regarding the results of statistical surveys as an input for decisions provides some guidelines in adjusting the calendar of data release to the users’ needs. In general, rational agents would appreciate less accurate data in advance instead of delayed perfect statistics, and the “impatience” of agents depends mainly on their capacity to make reliable early estimates of the relevant variables autonomously. In fact, provisional data assumedly improve their decisions only if data are capable of enhancing their own estimates and forecasts. Otherwise, rational agents would be better off continuing to base their decisions on their extrapolations. It follows that the size of forecast errors should be an important benchmark for statistical institutes in deciding when data should be released, taking into account the forecasting capability of “representative” data users, including government and professional users. As a consequence, regular surveys of users’ nowcasts could be helpful to enhance current release calendars.

The real data simulation experiment presented in Section 4 shows how the proposed approach may help to improve the current dissemination calendar of quarterly Euro area GDP. In particular, “flash” estimates seem just a little bit more accurate than naïve users’ forecasts made during the reference quarter, thus earlier (and coarser) releases would very likely be appreciated by users since such data could improve their nowcasts. Contrarily, the intermediate release of data 65 days after the end of the reference quarter apparently are less informative on the current dynamics of GDP,
since their accuracy does not beat the nowcasts already based on flash estimates solely. Of course, the breakdown of data provided by the second release is almost certainly valuable. In any case, statistical offices should balance such suggestions with the cost of producing more estimates and their institutional duties.

Further support for statistical agencies disseminating preliminary results of their surveys comes from the fact that rational agents often balance the cost of making a decision based on inaccurate data with the cost of delaying their decisions. If timing is crucial in making a decision, even very noisy and inaccurate preliminary data would be appreciated under most circumstances. However, according to the approach sketched above, designing a dissemination calendar requires first of all identifying the forecasting ability and the cost of postponing decisions of a “representative” data user. Notably, this conceptual framework seems fully consistent with the 11th principle of the European Statistics Code of Practice that states: “User satisfaction is monitored on a regular basis and is systematically followed up”, as well as with the 13th principle that provides that “Preliminary results of acceptable aggregate accuracy can be released when considered useful.”

In principle, the release of preliminary and final statistical data could be adapted dynamically to the possible changes of the accuracy of nowcasts, the variance of sample estimates and the cost of delaying decisions. Since predictions hopefully improve over time, also the publication of preliminary estimates from incomplete samples should be anticipated progressively. Furthermore, even less accurate statistical data could be appreciated by users just when their forecasts become more uncertain, for instance about the time of the turning points of the business cycle. Nevertheless, such a flexible dissemination policy could not comply with the commitment of statistical offices to follow a fully predictable strategy in order to strengthen their credibility and independence. Also data “inflation” could impair users, raising their search costs. Nevertheless, there is still room for flexibility in data release, provided that “Statistical release dates and times are pre-announced”, as stated by the 6th principle of the European Statistics Code of Practice, and “Divergence from the dissemination time schedule is publicised in advance, explained and a new release date set”, as pointed out by the 13th principle.

The comparison of users’ estimates vs. official preliminary sample estimates may also help official statisticians to decide the timing for the dissemination of disaggregated data. In fact, agents who need a given breakdown of data to make a decision, e.g. at $N$ “digits” level of the NACE classification of economic activity, necessarily compare the loss associated to the use of preliminary survey results at $N$ digit level, say $S^*(N)$, to the loss of using some model based estimation which exploits only data already available, e.g. data broken down at $N-n$ digits, say $F^*(N-n)$ already published. Thus, at time $t+h$, statistical data disaggregated at level $N$ would be long-awaited for by
agents only if $F^*(N-n) \geq S^*(N)$, otherwise users would be better off if statistical agencies had released earlier data disaggregated at level $N-n$ instead, that improve users extrapolations.

Further refinements of the approach presented in this paper and many more simulation experiments are required before implementing these concepts in official statistics. In particular, the cost of delaying decisions should be quantified to be compared to the loss related to the inaccuracy of data utilised in the decision process. Also, the advantages of model-based preliminary estimates directly released by the statistical offices, exploiting also internal and confidential information sources, should be explored, although this practice is often criticized by those defending a strict separation between official statistics and forecasting. Finally, an extensive analysis of releases of other statistical indicators is required. In any case, the suitability of releasing earlier preliminary data should be balanced with other considerations sketched above, mainly concerning the institutional role of statistical offices, and the cost incurred by users to collect and elaborate more information.

REFERENCES


European Communities (1999), Handbook on quarterly national accounts, Luxembourg.


