

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

*Enrico D'Elia*<sup>1</sup>

Economic agents are aware of incurring a loss in basing their decisions on their own extrapolations instead of on sound statistical data, but this loss may be smaller than the one related to waiting for the dissemination of the final data. Broad guidelines on deciding when statistical offices should release preliminary and final estimates of the key statistics may come from comparing the loss attached to users' predictions with the loss associated to possible preliminary estimates from incomplete samples. Furthermore, the cost of delaying decisions 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 requires some adjustment. In particular, actual “flash” estimates could be anticipated, while some later intermediate releases are likely less informative for the users.

*Key words:* Accuracy; data dissemination; Eurozone GDP; forecast; preliminary estimates; timeliness.

### 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 the same topic in depth 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 decision makers than the corresponding final releases. Economic agents also form their informed predictions on the relevant variables while waiting for official data releases. The main aim of this article 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

<sup>1</sup>Italian Ministry of Economy and Finance, via XX Settembre, 97, Rome 00187 and ISTAT, via Cesare Balbo, 16, Rome 00186, Italy. Email: [delia.enrico@gmail.com](mailto:delia.enrico@gmail.com)

**Acknowledgments:** The views expressed in this article are those of the author and do not necessarily reflect views at the Italian Ministry of Economy and Finance and ISTAT. The author gratefully acknowledges the valuable suggestions and criticisms made by 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.

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 of waiting for the dissemination of the final results of the pertinent statistical surveys before choosing between alternative strategies. In particular, the timing is important in most decision processes concerning investment, consumption, and price setting. 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.

Data users are perfectly aware that the final results of statistical surveys are more accurate than forecasts, nowcasts and early estimates. In principle, the 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, since 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 under “normal” conditions when no major shocks hit the economy or 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 and is subject to small sample bias 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 that it could be very costly. 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 it is hoped that the current data release calendar 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, primarily on their technical capability and knowledge.

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. Moreover, users' extrapolations hardly beat preliminary survey results when a major shock hits the economy, inevitably making model-based predictions 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 occasionally decreases.

Eurozone GDP estimates, analysed in the next sections, derive 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 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 conditions that no large shock perturbs the economy and that 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 adapt to the characteristics of some "representative" user of statistical data, endowed with a given capability and needs. Thus it is crucial to acknowledge that statistical offices must serve different users, including legislators and governmental agencies. 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 based 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 article.

## 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  $m_t$ . It is assumed that the “representative” economic agent has to base 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 after  $t$ .

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

- (a)  $f_{t+h} = E(m_t | \Omega_{t+h})$  the subjective prediction produced by exploiting the information set  $\Omega_{t+h}$ ;
- (b)  $s_{t+h}$  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 nonsample information. Although the Bayesian approach has many theoretical advantages, it is seldom used to improve sample estimates directly. 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 \(2011\)](#). 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 nonresponses (see [Graham et al. 2009](#)), to reduce the disclosure risk in the dissemination of individual data (see [Little et al. 2004](#)), to match the units of different surveys statistically (see [D’Orazio et al. 2006](#)), but not to improve preliminary estimates directly.

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

$$x_{i,t} = f_{t+h} + v_{t+h} + e_{i,t} \quad (1)$$

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 the predictability of the relevant events systematically. 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 cases the time invariance of  $\phi_h^2$  does not hold, while the variance of  $e_{i,t}$  does not necessarily change.

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 estimate of  $m_t$  at time  $t + h$  is

$$S_{t+h} = \sum_{i=1}^{M_{t+h}} w_{i,t+h} x_{i,t}, \quad (2)$$

where the weights  $w_{i,t+h}$  are such that  $\sum_{i=1}^{M_{t+h}} w_{i,t+h} = 1$  for each  $t + h$ . Under the previous assumptions on  $e_{i,t}$  in (1) and on the random collection of data, the average  $E(s_{t+h})$  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  $s_{t+h}$  is

$$\sigma_h = \frac{\sigma}{\sqrt{M_{t+h}}} \quad (3)$$

in the simplest case of equally weighted observations.

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

$$\sigma_h = \frac{\sigma}{\sqrt{M_0 + M_{t+h}}} \leq \frac{\sigma}{\sqrt{M_{t+h}}} \quad (4)$$

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 at its maximum when  $M_{t+h}$  equals zero, and is almost certainly large, unless the confidence of the statistical office in its priors is implausibly strong. In any case, during the survey,  $M_{t+h}$  is a nondecreasing 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 under “normal” conditions, in which data collected when  $h$  takes some special values are not systematically biased and volatile. Note that this is not the case when most influential units are surveyed just at the beginning and the end of the data collection process, for instance because some units are able to provide the data only according to a special calendar (e.g.: just after the balance sheets or periodic reports 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 nondecreasing 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 of a variable could be almost negligible in some cases, and potentially harmful in others. For instance, estimating the level and dynamics of GDP correctly is very important when deciding investment, but not export strategies. Nevertheless, the formal properties of  $L(\sigma_h)$  utilised 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 in particular 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 equally adverse to positive and negative estimation errors, in contrast to what [Granger and Pesaran \(2000\)](#) argue. However, 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 imbalances, due to larger social expenditure and lower tax revenues. Also, most firms acting in a competitive market fear overestimating the potential market much more than underestimating it, since overestimation calls for unduly large investment and related financial costs. By contrast, in oligopolistic markets, plants could be oversized intentionally to prevent the entry of possible competitors, so that entrepreneurs would be more averse 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 a small size of errors.

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  $h$ ,  $S(h)$  may also 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, or when information on the most important units can be collected. 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 nondecreasing function of  $\sigma_h$  implies that  $\frac{dS(h)}{dh}$  shares the sign of  $\frac{d\sigma_h}{dh}$ , apart from possible discontinuities. For instance, [Table 1](#) and [Figure 1](#) 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 industrial structure and the heterogeneity among firms' performances was completely different. 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 makes 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, namely when news available at time  $t+h$  prevails 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 the available information because collecting and elaborating it exceeds the profit expected from further improving their decisions. 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.

Like  $\sigma_h$ ,  $\phi_h$  can also 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, one can define  $F(h) = L(\phi_h)$ . Thus  $F(h)$ , similarly to  $S(h)$ , can be considered a nondecreasing 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

$$E(\sigma_h^2) = E \left[ \frac{1}{M_{t+h}} \sum_{i=1}^{M_{t+h}} (x_{i,t} - f_{t+h})^2 \right] - \phi_h^2 \quad (5)$$

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

As noted above, rational agents are assumed to be able to make forecasts even before data collection has begun, when  $\sigma_h^2$  is virtually infinite, so that  $\phi_h^2 < \sigma_h^2$  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 implausibly the latter would have only a little value for the representative user. Excluding the latter implausible case, the assumptions that  $\phi_h^2 < \sigma_h^2$  for  $h \leq 0$  and  $\frac{d\sigma_h}{dh} \leq \frac{d\phi_h}{dh}$  (disregarding possible singular discontinuities) subsequently 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  $\phi_h^2$  and  $\sigma_h^2$  are nonincreasing functions of  $h$ . Furthermore  $h_0$  could be very large, so that nowcasts could retain 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 incorporate even the very inaccurate preliminary data released before  $t + h_0$  into their nowcasts. 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



$h = h_I$ . The size of the shift can be ideally measured carrying out a survey among the users, or by assuming some nowcasting model, as in the next sections. 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 in the meantime continue to base their decision on their own past nowcasts even after the dissemination of official data. As a consequence, comparing the functions  $\phi_h$  and  $\sigma_h$  after each data release may provide operative guidelines for refining 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 of sample estimates is large. On the other hand, intermediate preliminary estimates that do not improve  $\phi_h$  sufficiently should be avoided, since they are very likely costly for the statistical offices and less appreciated by the users. It is worth noting that  $\phi_h$  and  $\sigma_h$  can be compared even 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 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 provided the best nowcast by applying efficient model-based estimators to the collected data, so that users' forecasts could hardly be better than 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 Särndal and Lundström (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 greatly among individuals. By contrast, the yield of a homogeneous set of equities can hardly be predicted, even if they have the same market price.

As assumed cautiously above, let the forecast accuracy improve over time less than  $\frac{\sigma_h}{\sigma_H}$ , namely 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

$$\sqrt{\frac{M_{t+H}}{M_{t+h}}} \geq \frac{\sigma_H}{\phi_h}. \quad (6)$$

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  $\frac{M_{t+h}}{M_{t+H}}$  that makes valuable the publication of preliminary results may be unexpectedly large, even when the prediction accuracy is quite poor compared to the 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 based on inaccurate data, often economic agents also have to 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 is able to serve the most profitable segment of demand, while the followers have to make do with supplying only the others. Furthermore, purchasing and investment decisions are usually supposed to have an optimal timing, mainly related to economic fluctuations. [Winston \(2008\)](#) provides 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 base 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 grow so fast 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  $\frac{dD}{dh} \geq 0$ .

As already noted in Section 2, both the accuracy of nowcasts and surveys may vary discontinuously over time, because most valuable data and information are often gathered only at specific points in time, and these points are often unpredictable, in particular for administrative sources. However, here  $F(h)$ ,  $S(h)$  and  $D(h)$  are assumed to be continuous functions of  $h$  only to make the problem more tractable analytically and show the role of the cost of delaying decisions in “normal” times.

In any case, if  $F(h)$  and  $S(h)$  cross for the first time at the delay  $h_0$ , as assumed in Section 2, the rational agents exploiting only predictions incur the minimum overall loss  $L_f$  at  $h_f$ ; thus  $L_f$  is approximately

$$L_f = (F(h_c) + D(h_0)) + (f' + d')(h_f - h_0) + (f'' + d'')(h_f - h_0)^2 \quad (7)$$

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

By contrast, users that base their decisions on the preliminary results of surveys face the minimum loss, say  $L_s$ ,  $h_s$  periods after the reference time, namely

$$L_s = (S(h_c) + D(h_c)) + (s' + d')(h_s - h_0) + (s'' + d'')(h_s - h_0)^2. \quad (8)$$

According to (7) and (8), the losses  $L_f$  and  $L_s$  achieve their minima when

$$h_f = h_0 - \frac{1}{2} \frac{f' + d'}{f'' + d''} \quad (9)$$

and

$$h_s = h_0 - \frac{1}{2} \frac{s' + d'}{s'' + d''} \quad (10)$$

that is when

$$L_f = (F(h_0) + D(h_0)) - \frac{1}{4} \frac{(f' + d')^2}{f'' + d''} \quad (11)$$

and

$$L_s = (S(h_0) + D(h_0)) - \frac{1}{4} \frac{(s' + d')^2}{s'' + d''}. \quad (12)$$

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), implies

$$\frac{(s' + d')^2}{s'' + d''} \geq \frac{(f' + d')^2}{f'' + d''}. \quad (13)$$

(13) entails that the decision makers would be better off basing their decisions on their own predictions even when the accuracy of nowcasts improves over time only very slowly, and much slower than the results of surveys, namely when  $f' \equiv 0$  and  $s'' \geq f''$  hold, so that (13) reads

$$s' \geq -2d' \quad (14)$$

namely if the marginal improvement of survey accuracy (i.e.  $-s'$ ) does not exceed twice the loss attached to postponing decisions by one unit of time more (i.e.  $d'$ ). Notably, the condition (14) derives from hypotheses that are very unfavourable to the use of forecasts and are less likely to occur in the real world. In any case, the inequality (14) fully confirms the assumption that agents prefer basing their decisions on predictions when the cost of delay increases very fast and the expected error size of surveys does not decrease too quickly over time. In the real world, nowcasts could improve quite fast, while the preliminary results of some survey may not. Thus the scope for utilising nowcasts is arguably even larger.

It is worth noting 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  $\phi_h$ , in contrast to  $S(h)$  and  $F(h)$ , thus the condition (14) strictly depends on the specific decision problem faced by economic agents. As a consequence, this factor cannot be 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 EC Regulation N° 1392/2007. However, the statistical offices of the member states and Eurostat tend to provide data in an even timelier manner 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, named “flash estimate” and consisting 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 lead to 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 eleven revisions during the subsequent twelve 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 (European Communities (1999) reports the methodologies and best practices for estimating quarterly national accounts in Europe). Thus quarterly GDP vintages almost certainly improve their accuracy over time much faster than a sequence of pure non-Bayesian estimates from incomplete samples such as that considered in Section 2 for illustrative purposes. Thanks to the mass of nonsampling information embodied in each release of data, the  $\sigma_h$  function associated to the GDP vintages can be expected to decrease 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, namely the 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 twelve countries of the Eurozone are collected and published regularly on the Eurostat website, starting from the rate of 2003Q1 (the revision triangle can be downloaded from [http://epp.eurostat.ec.europa.eu/portal/page/portal/national\\_accounts/methodology/quarterly\\_accounts](http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/methodology/quarterly_accounts) in Excel format). 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 six to nine months after the reference quarter, but in principle GDP can be revised many times for about three 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; accordingly, 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 third-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 article 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-day benchmark estimated on the sample 2003Q1 to

Table 1. The accuracy of preliminary estimates of GDP

Dissemination delay	Average error		RMSE		5th centile		95th centile	
	Full sample	Until 2008Q2	Full sample	Until 2008Q2	Full sample	Until 2008Q2	Full sample	Until 2008Q2
45 days	− 0.024	− 0.052	0.209	0.179	− 0.336	− 0.336	0.268	0.268
65 days	− 0.031	− 0.058	0.159	0.132	− 0.246	− 0.246	0.233	0.096
100 days	− 0.022	− 0.041	0.137	0.123	− 0.224	− 0.224	0.192	0.143
101–200 days	− 0.022	− 0.036	0.098	0.089	− 0.136	− 0.197	0.157	0.092
201–250 days	− 0.018	− 0.026	0.073	0.071	− 0.115	− 0.148	0.105	0.070
251–350 days	− 0.011	− 0.014	0.040	0.040	− 0.071	− 0.071	0.049	0.058

2011Q4 and on the pre-crisis subsample ranging from 2003Q1 to 2008Q2. The latter is part of a period often called the Great Moderation, because business cycle fluctuations, and average growth rate of GDP, 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 since World War II, and has provided many surprises for forecasters and statisticians.

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 a pure statistical point of view, possibly because the statistical offices are usually more concerned with overestimating GDP growth rates rather than with revising the data upward during the following years. Strikingly, this evidence was even stronger before the last economic crisis, supporting the view that the accuracy of official estimates has not been influenced overly 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 nine cases out of ten, 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 day releases, and the analysis of the pre-crisis period reveals similar results.

The same evidence is confirmed by the nonparametric estimate of the function  $\sigma_h$  reported in Figure 1, even again taking into account that few degrees of freedom are available particularly for the estimation on longer dissemination delays. The local second-degree

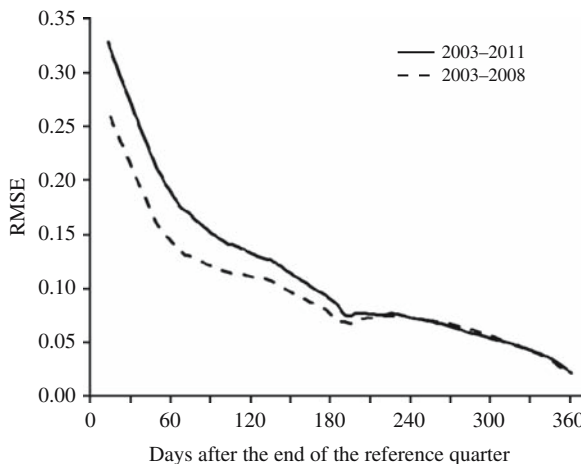


Fig. 1. The accuracy of quarterly GDP preliminary estimates

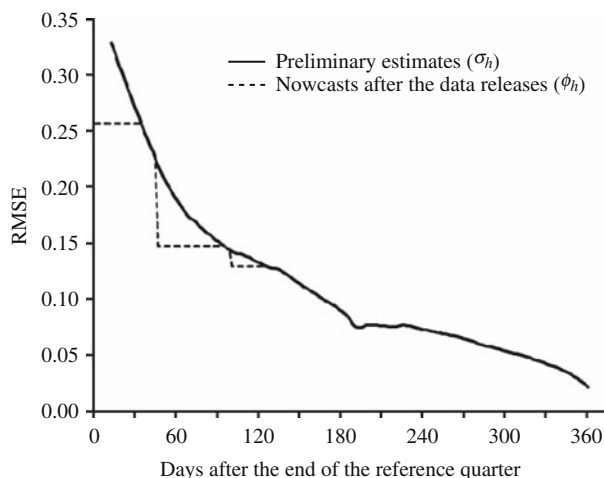


Fig. 2. Comparing the accuracy of nowcasts and preliminary estimates

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 a continuity of interpolated functions that could be unrealistic, as argued in Section 2.

Furthermore, according to nonparametric 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, almost 50% larger than the actual RMSE of flash estimates. Nevertheless, this value is likely underestimated, since it definitively comes from a purely 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.

Most results found comparing GDP revisions to their 400-day benchmarks are confirmed by considering the three-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, about 80% more than the RMSE computed versus the 400-day benchmark. The latter benchmark is still subject to revisions, the average size of which is 0.252 percentage points during the next 600 days. The nonparametric 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 nonparametric statistics for the three-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, and so on. 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 Barhouni 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 et al. (2008). In those papers, the RMSE of nowcasts based on real-time information likely available to data users ranges from 0.2 to 0.6 percentage points, with a gain of using additional information peaking as high as 40% of naïve predictions. Jansen et al. (2012) also estimated that consensus forecasts 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  $Y_t^*$  on the yearly growth rate of GDP made before the end of the reference quarter  $t$  derives from the simple AR model

$$Y_t^* = c_{t,v} + a_{t,v}Y_{t-1,v} + u_{t,v} \quad (15)$$

where  $c_{t,v}$ , and  $a_{t,v}$  are parameters estimated by using only the latest data available at time  $t-1$ , not including  $Y_t$ ;  $Y_{t-1,v}$  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 ten) would make it unfeasible to use other proper estimation methods. This practice creates 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.

When a new release of GDP figures, say  $Y_{t,v}$ , is published, users can improve their nowcast of the GDP growth by also taking into account 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

$$Y^*_{t,v} = c_{t,v} + a_{t,0,v-l}Y_{t,v-l} + a_{t,l,v-l}Y_{t-l,v-l} + a_{t,l,v}Y_{t-l,v} + d_{t,v}D_t + v_{t,v}, \tag{16}$$

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 ever 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 also reported by [Fixler and Grimm \(2006\)](#) for the US GDP, and by [Frale and Raponi \(2012\)](#) for the case of Italy.

The main results of estimating the models (15) and (16) from 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, the coefficients of  $Y_{t,v-1}$  are 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-day releases. By contrast, other regressors are not statistically significant: This result was expected for the dummy variable, which 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

Table 2. The main results of regressions used to simulate users’ nowcasts and forecasts (the statistics refer only to the longest sample available for each model)

Regressors	Forecast one quarter ahead	Data release		
		45 days	65 days	100 days
$Y_{t,v-1}$		1.059 (0.014)	1.033 (0.010)	1.030 (0.011)
$Y_{t-1,v-1}$		-0.044 (0.155)	0.297 (0.185)	0.013 (0.200)
$Y_{t-1,v}$	0.895 (0.079)	-0.008 (0.156)	-0.313 (0.187)	-0.039 (0.205)
Dummy variable		0.018 (0.074)	-0.002 (0.050)	0.174 (0.549)
Constant	0.133 (0.199)	-0.011 (0.014)	-0.015 (0.010)	-0.005 (0.010)
Adjusted $R^2$	0.795	0.999	0.999	0.999
RMSE	1.036	0.072	0.048	0.051

Standard error of estimates in parentheses.

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.

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 three percentage points below the true value), the predictions made before the end of the reference quarter are unexpectedly accurate, although they are intentionally naïve and 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 merits further attention, since the simulation period comprises the data on the last global crisis, when large unexpected shocks hit the European economy and a “double dip”, including three 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 5th and 95th 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 nonparametric interpolation of  $\sigma_h$  it emerges that even naïve users' predictions would be able to 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 estimates 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.

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 as “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 elaborating 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 greatly improve the accuracy of their adjusted nowcasts. The RMSE of nowcasts based on flash estimates is located 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

Table 3. *The accuracy of nowcasts and adjusted preliminary estimates of GDP*

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

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

<sup>(a)</sup>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.

substantially and intersects with the  $\sigma_h$  function when the dissemination delay is 95 days. In contrast, when the 65-day official estimates are released, users' adjusted nowcasts do not improve 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-day 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 become "rationally inattentive" as argued by [Sims \(2003\)](#), even if the accuracy of later official data releases of GDP is expected to increase. In any case, evidence for longer dissemination delays could be influenced by the scarce degrees of freedom available for estimation.

To summarise, 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 a 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 overly long before making their decisions.

A parallel simulation exercise carried out on data revisions versus the three-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 differs almost by 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

By regarding the results of statistical surveys as an input for decisions, we are able to provide some guidelines in adjusting the calendar to users' needs for data release.

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 are assumed to improve agents’ decisions only if the 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 in enhancing 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 only slightly 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. By contrast, the intermediate release of data 65 days after the end of the reference quarter apparently is less informative on the current dynamics of GDP, since the data’s accuracy does not surpass the nowcasts already based solely on flash estimates. 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. The empirical evidence presented in Section 4 also suggests that there is only little scope for users to wait for definitive data published after three years before making their decisions, since the revisions made beyond 400 days after the reference quarter are generally small, apart from general methodological changes that appear virtually independent compared to the first revisions.

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 the identification of the forecasting ability of and the cost of postponing decisions to 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, the publication of preliminary estimates from incomplete samples should be anticipated progressively. Furthermore, even less accurate statistical data about, for instance, the turning points of the business cycle could be appreciated by users when their forecasts become more uncertain. Nevertheless, such a flexible dissemination policy would not comply with statistical offices’ commitment to following a fully predictable

strategy in order to strengthen their credibility and independence. Moreover, 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 preannounced”, as stated by the 6th principle of the European Statistics Code of Practice, and “[d]ivergence 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 versus 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, for example 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, such as data broken down at  $N - n$  digits, say  $F^*(N - n)$ . Thus, at time  $t + h$ , statistical data disaggregated at level  $N$  would be long-awaited 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. However, more research is probably needed to thoroughly investigate the issue of how and when preliminary disaggregated data should be disseminated.

Further refinements of the approach presented in this article 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. Furthermore, the advantages of model-based preliminary estimates directly released by the statistical offices, also exploiting 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 in collecting and elaborating more information.

## 6. References

- Altavilla, C. and M. Ciccarelli. 2007. “Information Combination and Forecast (st)ability. Evidence from Vintages of Time-Series Data.” *Working Paper Series ECB*, No. 864. Available at: <http://ideas.repec.org/e/pal73.html> (accessed July 31, 2014).
- Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler. 2011. “Short-Term Forecasts of Euro Area GDP Growth.” *The Econometrics Journal* 14: 25–44. DOI: <http://dx.doi.org/10.1111/j.1368-423X.2010.00328.x>.
- Barhoumi, K., S. Benk, R. Cristadoro, A. Reijer, P. Jakaitiene, P. Jelonek, and A. Rua. 2008. “Short-Term Forecasting of GDP Using Large Monthly Datasets: a Pseudo Real-Time Forecast Evaluation Exercise.” *Occasional Paper Series ECB*, No. 84. Available at: <http://ideas.repec.org/p/ecb/ecbops/20080084.html> (accessed July 31, 2014).

- Blanchard, O.J., J.P. L'Huillier, and G. Lorenzoni. 2009. "News, Noise, and Fluctuations: An Empirical Exploration." *NBER Working Paper*, No. w15015, Available at: <http://www.nber.org/papers/w15015> (accessed July 31, 2014).
- Clemen, R. 1989. "Combining Forecasts: a Review and Annotated Bibliography." *International Journal of Forecasting* 5: 559–583. DOI: [http://dx.doi.org/10.1016/0169-2070\(89\)90012-5](http://dx.doi.org/10.1016/0169-2070(89)90012-5).
- D'Orazio, M., M. Di Zio, and M. Scanu. 2006. *Statistical Matching: Theory and Practice*. New York: Wiley.
- Diron, M. 2008. "Short-Term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-Time Performance Based on Vintage Data." *Journal of Forecasting* 27: 371–390. DOI: <http://dx.doi.org/10.1002/for.1067>.
- European Central Bank 2009. "Revisions to GDP Estimates in the Euro Area." *Monthly Bulletin* 4: 85–90. Available at: <https://www.ecb.europa.eu/pub/pdf/mobu/mb200904en.pdf> (accessed July 31, 2014).
- European Communities 1999. *Handbook on Quarterly National Accounts*, Luxembourg: Office for Official Publications of the European Communities. Available at: [http://epp.eurostat.ec.europa.eu/cache/ITY\\_OFFPUB/CA-22-99-781/EN/CA-22-99-781-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/CA-22-99-781/EN/CA-22-99-781-EN.PDF) (accessed July 31, 2014).
- European Statistics Code of Practice 2011. Available at: [http://epp.eurostat.ec.europa.eu/cache/ITY\\_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF).
- Fan, J. 1992. "Design-Adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87: 998–1004. DOI: <http://dx.doi.org/10.2307/2290637>.
- Fan, J. and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.
- Fixler, D.J. and B.T. Grimm. 2006. "GDP Estimates: Rationality Tests and Turning Point Performance." *Journal of Productivity Analysis* 25: 213–229. DOI: <http://dx.doi.org/10.1007/s11123-006-7640-x>.
- Frale, C. and V. Raponi. 2012. "Revisions in Official Data and Forecasting." *Working Papers of Dipartimento del Tesoro*, No. 3.
- Frale, C., M. Marcellino, G.L. Mazzi, and T. Proietti. 2011. "EUROMIND: A Monthly Indicator of the Euro Area Economic Conditions." *Journal of the Royal Statistical Society: Series A* 174: 439–470. DOI: <http://dx.doi.org/10.1111/j.1467-985X.2010.00675.x>.
- Giannone, D., L. Reichlin, and D. Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55: 665–676. DOI: <http://dx.doi.org/10.1016/j.jmoneco.2008.05.010>.
- Graham, P., J. Young, and R. Penny. 2009. "Multiply Imputed Synthetic Data: Evaluation of Hierarchical Bayesian Imputation Models." *Journal of Official Statistics* 25: 245–268.
- Granger, C.W.J. and M.J. Machina. 2006. "Forecasting and Decision Theory." In *Handbook of Economic Forecasting*, edited by G. Elliott, C.W.J. Granger, and A. Timmermann. Amsterdam: Elsevier.
- Granger, C.W.J. and M.H. Pesaran. 2000. "Economic and Statistical Measures of Forecast Accuracy." *Journal of Forecasting* 19: 537–560. DOI: [http://dx.doi.org/10.1002/1099-131X\(200012\)19:7<537:AID-FOR769>3.0.CO;2-G](http://dx.doi.org/10.1002/1099-131X(200012)19:7<537:AID-FOR769>3.0.CO;2-G).

- Jansen, W.J., X. Jin, and J. de Winter. 2012. "Forecasting and Nowcasting Real GDP: Comparing Statistical Models and Subjective Forecasts." *De Nederlandsche Bank Working Paper*, No. 365. Available at: [http://www.dnb.nl/en/binaries/Working%20Paper%20365\\_tcm47-283164.pdf](http://www.dnb.nl/en/binaries/Working%20Paper%20365_tcm47-283164.pdf) (accessed July 31, 2014).
- Little, R.J.A. 2012. "Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics." *Journal of Official Statistics* 28: 309–334.
- Little, R.J.A., F. Liu, and T.E. Raghunathan. 2004. "Statistical Disclosure Techniques Based on Multiple Imputation." In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, edited by A. Gelman and X.L. Meng, 141–152. New York: John Wiley & Sons.
- Pain, N. and F. Sédillot. 2005. "Indicator Models of Real GDP Growth in the Major OECD Economies." *OECD Economic Studies*, No. 40. Available at: <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=4a74b653-6721-45c4-897d-9aa24c0c2037%40sessionmgr113&vid=2&hid=128>. (accessed July 31, 2014).
- Särndal, C.-E. and S. Lundström. 2005. *Estimation in Surveys With Nonresponse*. New York: John Wiley & Sons.
- Sims, C.A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50: 665–690. DOI:[http://dx.doi.org/10.1016/S0304-3932\(03\)00029-1](http://dx.doi.org/10.1016/S0304-3932(03)00029-1).
- UNSTAT, 2009. "International Seminar on Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends." Available at: <http://unstats.un.org/unsd/nationalaccount/workshops/2009/ottawa> (accessed July 31, 2014).
- Winston, G.C. 2008. *The Timing of Economic Activities*. Cambridge: Cambridge University Press.
- Yang, Y. and H. Zou. 2004. "Combining Time Series Models for Forecasting." *International Journal of Forecasting* 20: 69–84. DOI: [http://dx.doi.org/10.1016/S0169-2070\(03\)00004-9](http://dx.doi.org/10.1016/S0169-2070(03)00004-9).

Received April 2012

Revised March 2014

Accepted March 2014