DISCUSSION PAPER SERIES

No. 5302
CEPR/EABCN No. 24/2005

SHORT-RUN ITALIAN GDP FORECASTING AND REAL-TIME DATA

Roberto Golinelli and Giuseppe Parigi

INTERNATIONAL MACROECONOMICS

€ABCN

Euro Area Business Cycle Network

www.eabcn.org

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5302.asp**

ISSN 0265-8003

SHORT-RUN ITALIAN GDP FORECASTING AND REAL-TIME DATA

Roberto Golinelli, Università di Bologna
Giuseppe Parigi, Banca d'Italia

Discussion Paper No. 5302
CEPR/EABCN No. 24/2005

October 2005

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Roberto Golinelli and Giuseppe Parigi

CEPR Discussion Paper No. 5302
CEPR/EABCN No. 24/2005

October 2005

ABSTRACT

Short-Run Italian GDP Forecasting and Real-Time Data*

National accounts statistics undergo a process of revisions over time because of the accumulation of information and, less frequently, of deeper changes, as new definitions, new methodologies etc. are implemented. In this paper we try to characterise the revision process of the data of Italian GDP as published by the national statistical office (ISTAT) in the stream of the noise models literature. The analysis shows that this task can be better accomplished by concentrating on the growth rates of the data instead of the levels. Another issue tackled in the paper concerns the informative content of the preliminary releases *vis* a *vis* an intermediate vintage supposed to embody all statistical information (or no longer revisable as far as purely statistical changes are concerned) and the latest vintage of the data, supposed to be the definitive one. The analysis of the news models in differences is based on the comparison of the forecasting performance of the preliminary releases with that of a number of one step ahead forecasts computed from alternative models, ranging from very simple univariate to multivariate specifications based on indicators (bridge models). Results show that, for the intermediate vintage, the preliminary version is the better forecast, while the latest vintage, which embodies statistical as well as definitional revisions, may be better characterised by considering both the preliminary version and the bridge models forecasts.

JEL Classification: C22, C53, C82 and E10
Keywords: consistent vintages, predictions of 'actual' GDP, preliminary GDP forecasting and real-time data set for Italian GDP

Roberto Golinelli
Dept of Economics
University of Bologna
Strada Maggiore 45
40125 Bologna
ITALY

Tel: (39 051) 209 2638
Fax:(39 051) 209 2664
Email: golinell@spbo.unibo.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158399

Giuseppe Parigi
Servizio Studi
Banca d'Italia
Via Nazionale
91-00184 Roma
ITALY

Tel: (39 06) 4792 3085
Fax:(39 06) 4792 3720
Email:
giuseppe.parigi@bancaditalia.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135846

*The publication of this Paper is funded by the Euro Area Business Cycle Network (www.eabcn.org). This Network provides a forum for the better understanding of the euro area business cycle, linking academic researchers and researchers in central banks and other policy institutions involved in the empirical analysis of the euro area business cycle.

Submitted 19 September 2005

SHORT-RUN ITALIAN GDP FORECASTING AND REAL-TIME DATA

by Roberto Golinelli* and Giuseppe Parigi**

Abstract

National accounts statistics undergo a process of revisions over time because of the accumulation of information and, less frequently, of deeper changes, as new definitions, new methodologies etc. are implemented. In this paper we try to characterise the revision process of the data of Italian GDP as published by the national statistical office (ISTAT) in the stream of the noise models literature. The analysis shows that this task can be better accomplished by concentrating on the growth rates of the data instead of the levels. Another issue tackled in the paper concerns the informative content of the preliminary releases *vis a vis* an intermediate vintage supposed to embody all statistical information (or no longer revisable as far as purely statistical changes are concerned) and the latest vintage of the data, supposed to be the definitive one. The analysis of the news models in differences is based on the comparison of the forecasting performance of the preliminary releases with that of a number of one step-ahead forecasts computed from alternative models, ranging from very simple univariate to multivariate specifications based on indicators (bridge models). Results show that, for the intermediate vintage, the preliminary version is the better forecast, while the latest vintage, which embodies statistical as well as definitional revisions, may be better characterised by considering both the preliminary version and the bridge models forecasts.

Index

This version: July 30th 2005

JEL classification: C53, C22, C82, E010

Keywords: Real-time data set for Italian GDP, consistent vintages, preliminary GDP forecasting, predictions of "actual" GDP.

* Department of Economics, University of Bologna, Italy (golinell@spbo.unibo.it)

** Research Department, Bank of Italy (giuseppe.parigi@bancaditalia.it)

1. Introduction[1]

Decision makers in different branches of the economy (business, governments, central banks, traders in financial markets, etc.) operate in real-time, and their choices are based on the early understanding of the state of the economic activity (generally measured by real GDP). When a new GDP series is released by statistical agencies, practitioners typically base their analyses of the short-run economic evolution only on these data and update all previous work correspondingly (both analyses and forecasts). Two broad assumptions lay at the roots of such practice: (*a*) when new measurements of GDP are issued, their growth rates embody all the relevant information (in the sense that levels *per se* do not carry useful insights); (*b*) though preliminary, first-published GDP growth rates deliver the only valid description of the actual economic situation: previous forecasts for that period are no longer informative and are discarded. This paper try to investigate the statistical appropriateness of these two hypotheses with reference to the Italian case but, in doing so, we introduce a methodological approach that can also be applied to other countries.

It is a matter of fact that statistical agencies usually and frequently revise their estimates of economic aggregates as new information becomes available. The first set of national accounts (NA, hereafter) estimates for a given period may be generally revised several times. Therefore, an assessment of the validity of hypotheses sub *(a)* and *(b)* should be made by using the whole historical information set of what has been published in real-time (*i.e.* the whole set of revisions).

Real-time analyses are not new in the literature: since the beginning of the 50s, many researchers have dealt with the implication of data revisions for forecasting (mainly for the US case). During the recent years the number of contributions has sharply increased also because of the availability of large - internet-downloadable - data sets (Dean Croushore and Tom Stark pioneered, in the late 90s, this practice with the US real-time data set posted at the Fed of Philadelphia web site).

The aim of this paper is twofold. Firstly, we want to assess the existence of valid relationships among different GDP provisional measurements, by taking into account both

[1] Paper presented at the CEPR-EABCN 5th Workshop: “Needed: A Real Time Database for the Euro-Area”, Brussels, National Bank of Belgium, 13-14 June 2005. We are grateful to Fabio Busetti, Tommaso Di Fonzo, Daniele Garbarino, David Hendry, Stefano Siviero, Giovanni Veronese and workshop participants for comments and suggestions. The views contained here are those of the authors only and do not necessarily reflect those of the institutions for which they work.

levels and growth rates. Such analyses imply an arbitrary definition of a new variable measuring GDP levels released over time. In short, we show that there is enough evidence (at least for the Italian case) supporting the statistical appropriateness of disregarding levels: both preliminary and intermediate versions of GDP are never cointegrated with the latest available (actual) GDP levels, supposed to be no longer revisable.

Secondly, conditional on previous findings, we interpret preliminary growth rates as predictions of the actual GDP data, and assess their forecasting ability with respect to a number of alternative models. Although a very high weight should be placed on preliminary data, we find that the prediction of the latest available Italian data could be improved by combining preliminary releases and one step-ahead forecasts.[2]

The paper is organised as follows. Section 2 describes main definitions used in this paper with the help of a schematic representation of a typical real-time data set. In particular, we distinguish between the definitions of "vintage" and "outturn", and discuss alternative ways to measure real-time levels. These definitions are then applied to the Italian GDP, and results from preliminary analyses of the data are presented. In Section 3 we describe the framework used in the real-time forecasting exercise (with random walk, ARIMA, leading indicators and bridge models), then we assess the ability of alternative one-quarter ahead model-based predictions in forecasting the preliminary GDP release. In Section 4 we compare the ability of both the first GDP release and model-based predictions in forecasting "actual" GDP data. In doing so, we also make tests of forecast encompassing in order to assess whether practitioners should either discard their forecasts as soon as preliminary GDP data are released, or combine the two. Finally, Section 5 concludes.

2. The real-time data set for Italy

2.1 Basic concepts and definitions

Statistical agencies often revise GDP data because of statistical and definitional changes (in our paper, as in most of the literature, we concentrate on quarterly data). Statistical changes stem from the availability of additional information as time elapses, and generally only concern the most recent quarters. Definitional changes (in the base year, and/or due to methodological reforms such as changes in classification) are more pervasive and

---

[2] After the seminal studies of Rettore and Trivellato (1986), and Bordignon and Trivellato (1989), the Italian case has been recently analysed by Faust *et al.* (2005).

occur at discrete times, say every four-eight years (depending on the country and the period) involving a retrospective change of the whole historical sample.

In this paper, we define each GDP series of the real-time data set in Table 1 as ${}^{o}y^{v}_{t}$ where o = 1, 2, ..., T+f is the *outturn* index; v = 1, 2, ..., f is the *vintage* index (f labels the final, *i.e.* latest available, vintage); t = 1, 2, ..., T+f is the *period* index. In the data matrix reported in Table 1, rows are periods, columns are vintages (*i.e.* y^{v} is the GDP time series over the period 1 to T+v published by the statistical agency in its issue v), and diagonals are outturns. The sequence of observations ${}^{1}y^{1}_{T+1}$ ${}^{1}y^{2}_{T+2}$ ${}^{1}y^{3}_{T+3}$ ... ${}^{1}y^{f}_{T+f}$ is the time series of the first outturn ${}^{1}y$ of GDP data that have never been revised, while the time series of the second outturn ${}^{2}y$ contains data that have all been revised once, and so on.[3] The structure of Table 1 reflects the way in which real-time data sets are usually stored.[4]

Table 1 here

The T^{th}-period row splits the data matrix in Table 1 into two parts: the first (T×f) sub-matrix merges all vintage series for a common period with no missing values. In other words, T is the latest possible period until which all vintages in the data set have values. In particular, the first vintage has only one quarter after T; the second vintage has two quarters and so on until the last vintage, the f-th one, which has f values. In the case of Italy as the official quarterly data start from 1970 Q1, the last quarter of the first vintage in our data set is 1988 Q1 so that T = 73 (in the same way, T = 74 for the USA and T = 25 for the UK). GDP levels of the same vintage v reflect the state-of-the-knowledge at T+v and refer to different stages of the revision process. The growth rates of the v^{th} vintage is given by:

$$ {}^{o}dy^{v}_{t} = 100 \times ({}^{o}y^{v}_{t} / {}^{o'}y^{v}_{t-1} - 1), \text{ where } o' = o+1, \text{ and } t = 2, 3, ..., T+v.$$

[3] There is no accepted set of definitions in the literature. With some *caveats*, "measurement" (Garratt and Vahey, 2004), "release" (Croushore and Stark, 2003, and Swanson and van Dijk, 2004), "announcement" (Mankiw *et al.*, 1984), "estimate" (Mankiw and Shapiro, 1986, Mork, 1987, and Roodenburg, 2004), "real-time vintage" (Koenig *et al.*, 2003), and "real-time data" (Dopke, 2004, and Clausen and Meier, 2003) are all synonyms of our "outturn" definition that we owe to Pain (1994). On the other hand, Patterson and Heravi (1991), Patterson (2000, 2002, 2003), Siklos (1996) and Busetti (2001) define as "vintage" what we define as "outturn".

[4] See the real time datasets for the US (http://www.phil.frb.org/econ/forecast/reaindex.html) and for the UK (http://www.bankofengland.co.uk/statistics/gdpdatabase/). The matrix representation of the real-time data set in Busetti (2001, Table 1) and in Patterson and Heravi (1991, Table 1) is defined exactly as the opposite of our Table 1: outturns are in columns, and vintages are in diagonals. Koenig *et al.* (2003, appendix) report a similar structure, but start from growth rates instead of levels for reasons that we describe in paragraph 2.3.

The GDP time series by vintage, in levels y^v or in quarterly growth rates dy^v, can be used as they are published, in modelling and in forecasting with the only *caveat* of their provisional nature.

The values on the diagonals in Table 1 come from different vintages (*e.g.* different base-years, classification, etc.) and it clearly makes no sense to compute their growth rates. In the real-time literature, this problem is avoided by building a matrix of growth rates (${}^{o}dy$) instead of levels[5]. In this way, the values on the diagonals are directly the growth rates computed consistently for each vintage. We define ${}^{o}dy$ as the o^{th} outturn growth rate *within* vintages (growth-within, hereafter); in this case the time series of the first outturn of GDP growth-within (${}^{1}dy_t$ for t = T+1, T+2, ..., T+f) is the sequence of the growth rates computed on the preliminary GDP series published by the statistical agency.[6]

In the real-time context the use of levels of a series (see paragraph 2.3), requires a preliminary transformation of the data in order to remove the heterogeneity effects induced by definitional changes. This can be done in two alternative ways: by using rescaling factors (*e.g.* normalising to a common base period or deflating values at current prices with common base-period price indexes); by applying the regression approach (see Patterson and Heravi, 1991). While the first method is clear and simple, in the second one ${}^{o}y^{v}_{t}$ should be modelled (*e.g.* how to account for integration-cointegration data properties) and its application in further econometric analyses raises problems of generated regressors.

The practice of rescaling series to a common base-period is needed for comparisons (rescaled data reflect cumulative growth rates since the base period; see for example, the country GDP volume index in 1995 base published by the IMF International Financial Statistics). However, in the real-time case, rescaling is fairly restrictive because it assumes that all vintages are consistent in the base period τ (*i.e.* vintages in τ are supposed to differ only because of different base years; when this is not true rescaling and vintage effects are confused, see Patterson and Heravi, 1991). The transformed series can be computed from the ratios (where $\tau = 1$ is assumed to be the beginning of the sample):

---

[5] Growth rates are originally used in Steckler (1967), Mankiw *et al.* (1984) and in Mankiw and Shapiro (1986); more recently, in Croushore and Stark (2003), Koenig *et al.* (2003), Swanson and van Dijk (2004) for the US; Castle and Ellis (2002), Garratt and Vahey (2004) for the UK; Roodenburg (2004) for the Netherlands; De la Rocque *et al.* (2004) for Brazil; Faust *et al.* (2005) for the G7 countries.

[6] When we collect the time series for the o^{th} outturn of GDP growth rates we assume that they are not affected by different definitions and methodological changes (see Mork, 1997); however, Garratt and Vahey (2004) and Swanson and van Dijk (2004) find non constancy and non linearity due to the revision process.

$$^{o}x^{v}_{t}, = {}^{o}y^{v}_{t} / {}^{o}y^{v}_{1} .$$

Since T is usually large, it seems reasonable to assume that, in the first period, the levels of various vintages differ only because of different base years; in effects, this is the case of the Italian, American and English data sets[7].

The normalisation could be applied to all ${}^{o}y^{v}_{t}$ data of Table 1, but for real-time data revisions analysis we focus only on the (f×f) upper triangular sub-matrix of ${}^{o}x^{v}_{t}$, where t=T+1, T+2, ..., T+f, and v=1, 2, ..., f. The corresponding growth rates may be defined as the time series of the o[th] outturn of growth rates *between* vintages (growth-between, hereafter):

$$^{o}dx^{v}_{t} = 100 \times ({}^{o}x^{v}_{t}/{}^{o}x^{v'}_{t-1} - 1), \text{ where } v' = v-1, \text{ and } t = T+2, T+3, ..., T+f-o+1.$$

As growth-between data are affected by short-run macroeconomic patterns as well as revisions, they are not used by practitioners, which usually prefer growth-within data for their analyses of the business cycle. So we propose a second definition of GDP levels, say ${}^{o}z^{v}_{t}$, which is consistent across vintages and preserves growth-within information:

$$^{o}z^{v}_{t} = {}^{o}z^{v}_{t-1} \times (1 + {}^{o}dy^{v}_{t}/100), \text{ where } t = T+2, T+3, ..., T+f-o+1.$$

In this case the time series of the levels of the o[th] outturn, ${}^{o}z$, can be computed by iterating the formula above starting from the initial condition ${}^{o}z^{v}_{T+1} = {}^{o}x^{v}_{T+1}$. It is easy to see that in T+1 ${}^{o}z^{v}_{t}$ levels are equal to those of ${}^{o}x^{v}_{t}$ and their growth rates coincide with growth-within data.

2.2 A preliminary inspection of data

Concepts and definitions of the previous paragraph can be applied to the data set of GDP vintages issued by the Italian Statistical Agency (ISTAT). The first vintage of our data-set (v=1) reports GDP data from 1970 Q1 to 1988 Q2, while the latest available vintage (v=f=67) reports data from 1970 Q1 to 2004 Q4. From 1988 to 2004, Italian NA experienced five benchmark revisions: base-year change from 1980 to 1985, involving 12 vintages; base-year change from 1985 to 1990, 31 vintages; preliminary changes from SEC 79 to SEC 95, and base-year change from 1990 to 1995, 43 vintages; complete retrospective SEC 95 data, from 1970 Q1, 51 vintages (since the 55[th] vintage GDP unit was proportionately transformed

[7] If GDP levels at current prices reflect the revisions of interest, rescaling may be accomplished by computing GDP at constant prices with a deflator expressed in the *same* base period for *all* vintages (see Patterson, 2002). With this method the normalised constant price GDP levels are conditional on the specific choice of the deflator and no longer match actual data.

into the euro). Finally, since the 60^{th} vintage GDP levels are adjusted for trading days. Table 2 summarises the main features of the Italian real-time data set[8].

Table 2 here

In paragraph 2.1 we suggested two alternative measurements of vintage-consistent GDP levels, both from published ${}^{o}y^{v}_{t}$ series: ${}^{o}x^{v}_{t}$ (obtained by rescaling) and ${}^{o}z^{v}_{t}$ (obtained by cumulating GDP growth-within). For simplicity, in Figure 1 only seven of such variables for our data set are reported: the first log-outturns of both x- and z-type data (${}^{1}x$, ${}^{1}z$), the corresponding intermediate fourth and eighth outturns (${}^{4}x$, ${}^{4}z$, and ${}^{8}x$, ${}^{8}z$) that should embody the main portion of the statistical revisions, and the finally revised data (*i.e.* the latest available vintage, x^{67}).

Figure 1 here

Different data transformations are reported in each row of graphs. The first row depicts the patterns of GDP log-levels, the second row both the differences among the first, the fourth and the eighth outturn log-levels and the latest available vintage (*i.e.* the revision with respect to the latest available data of the first, the fourth and the eighth outturn). The third row plots GDP growth-between and growth-within, along with the growth rates of the latest available vintage. Some preliminary findings can be summarised as follows.

1) Alternative GDP levels in the upper panel appears to be integrated stochastic processes, as suggested by a wide range of unit-root tests results (not reported). Since the first differences of all the series (in the lower panel) appear to be stationary, both outturns and latest available GDP log-levels are first-order integrated, I(1).

2) In the central panel of Figure 1, total revisions (measured by the log-difference between the first, the fourth and the eighth outturn and the latest available vintage) are very persistent over time: definitional revisions permanently move away actual GDP from z-outturn levels, but the fact that *log*(${}^{o}z/x^{67}$) levels converge to a constant, imply their convergence in growth rates.

[8] The revision process for the Italian quarterly national accounts data is described in Giovannini (1993), Di Fonzo *et al*. (2002), and Lupi and Peracchi (2003).

3) The variances of the first, the fourth and the eighth outturns growth rates are all very similar and unrelated with o (Figure 1, lower panel): the p-values of alternative tests for equality of their variances are about 20% (for x data) and 80% (for z data). However, the growth-within variances are significantly smaller than the growth-between ones, probably because of the presence of large outliers in x-outturns, mainly for the first release.

In general, findings 1 and 2 suggest that cointegration analysis should be based on alternative level definitions (*i.e.* both x and z data). Findings 2 and 3 show that relevant revisions makes both x- and z-outturn levels incompatible with the latest available series; growth-between (^{o}dx) data appear inconsistent over time with the official GDP figures.

2.3 The revision process of GDP data

The aim of this section is twofold: (i) to characterise the data-revision process for Italy with appropriate statistical analysis, and (ii) to suggest what kind of data (either levels or growth rates) is better to use when GDP data are released.

Following the seminal paper by Mankiw and Shapiro (1986), a large part of the real-time literature exploits growth-within data instead of levels[9]; levels however are still used by a different branch of the literature, where cointegration techniques are applied.[10] We merge these two approaches by setting a model where I(1) outturn levels and I(0) growth rates are both included in cointegrated VAR models. The analysis is applied to the Italian real-time data set described above, and is organised in subsequent steps, from univariate to multivariate, in order to cross-validate results of different approaches.

At the beginning the stationarity of both sequential and total revisions is evaluated according to the Elliott *et al.* (1996) test, respectively in the upper and in the lower panel of Table 3. Sequential revisions are defined as the logs of the ratio between two successive outturns; total revisions are the logs of the ratio between one outturn and the latest available vintage. If either sequential or total revisions are stationary, the involved outturn/latest available levels are (1, -1) cointegrated. The statistics for ^{o}x are in the third column; those for cumulated growth-within ^{o}z in the fifth column.

Table 3 here

[9] Recent contributions are Garratt and Vahey (2004), Swanson and van Dijk (2004), and Faust, *at al.* (2005).

[10] Patterson and Haravi (1991), Pain (1994), and Siklos (1996) test for Engle and Granger (1987) cointegration between quarterly NA outturns; Gallo and Marcellino (1999) and Patterson (2000, 2002, and 2003) apply the more general Johansen (1995) approach.

The outcomes of the tests are fairly clear: the sequential revisions for ${}^{o}x$ are often stationary, while those for ${}^{o}z$ are always integrated; total revisions are non stationary, as shown in the second row of Figure 1. In other terms, a cointegrated level relationship is only detected for the x (*i.e.* rescaled) log-levels released in the first two years. Latest available levels (x^{67}) do not cointegrate with outturn levels (both ${}^{o}x$ and ${}^{o}z$) because of benchmark revisions, which sometimes imply relevant changes in the measurement system. This is what basically motivates the use of growth rates by practitioners.

The same results can be obtained by considering the less restrictive Johansen (1995) approach for all the eight x-outturns together. In particular, results in the first row of Table 4 show that sequential revisions of x levels are stationary (similar results for the US industrial production index are in Patterson, 2000, and for the UK quarterly GNP in Patterson, 2002). Moreover, after a number of not rejected restrictions are imposed to the VAR (in the last column of Table 4) the eighth outturn ${}^{8}x$ appears to be the only weak exogenous variable of the system (the only permanent component of the first eight x-outturns, see Patterson, 2003). The statistical (sequential) revisions of x levels are only transitory because, during the first two years of releases, the accumulation of new information allows the Italian statistical agency (ISTAT) to reduce the noise embodied in the preliminary x-outturn levels.

Table 4 here

Results do not change if we omit from the VAR a number of x outturns, for example by keeping only the preliminary (the first) and two intermediate (the fourth and the eighth) outturns (see the second and the third row of Table 4). The latest available vintage x^{67} does not seem to play any role in the level relationships, thus confirming the results reported in the lower panel of Table 3. The inclusion of the latest available vintage does not increase the cointegration rank because these data embody a different stochastic trend from the one driving the first eight outturns.

If we make the same kind of experiment with the z-levels (that cumulate GDP growth-within) the absence of cointegration in the two systems is never rejected (fourth and fifth rows in Table 4) and corroborates the findings in Table 3. When we put together x- and z-outturns (preliminary and intermediate) and the latest available data (see the sixth row of Table 4), the inclusion of the z-outturn levels does not increase the rank (always equal to one) of the VAR (third row).

As a final remark, the analysis of the time series properties of the revisions of the Italian GDP data has shown that it is difficult to find a relationship between the levels of a series (however transformed to account for heterogeneity) and growth-within data. In other terms, growth rates of preliminary estimates, *i.e.* $\Delta log(^1z)$, seem to be not affected statistical revisions, *i.e.* $log(^1x/^8x)$, thus supporting the practice of concentrating on growth rates rather than levels.

Moreover, in all exercises described above the inclusion of the latest available vintage does not lead to any level relationship with other levels, essentially because benchmark revisions often embody deeper changes that may completely alter the properties of a series. In this context, it may be interesting to try an assessment about the informative content of one-step ahead GDP growth forecasts, along with that of preliminary releases, in order to predict the growth of fully revised data. If the news (or rational forecast) model is valid, "actual" (*i.e.* after that many revised data are released) GDP growth should be unpredictable by using any other information available at the time of the first outturn (see Mankiw *et al.*, 1984, Mankiw and Shapiro, 1986, and Faust *et al.*, 2005). Given that Busetti (2001) and Faust *et al.* (2005) clearly reject the news model for the Italian case, in the next Sections we will assess this topic in a real-time forecasting framework.

3. Model-based GDP forecasts in real-time

In this Section we describe the empirical framework to forecast GDP in real time, and to assess the ability of alternative econometric models to predict the first outturn, *i.e.* the GDP growth rate of the first release. Paragraph 3.1 reviews main methodological issues involved by the construction of our forecasting tools and reports main estimation results; paragraph 3.2 assesses our models forecasting performance in real-time for the period 1991-2004.

3.1 The empirical framework and in-sample analysis

In order to set up an empirical framework to mimic as close as possible the real-time forecasting activity, we need three basic ingredients: (1) a real-time data set representing the data availability at any given date in the past; (2) a number of models covering a wide range of the trade-off between simple and complex models; (3) some behavioural assumptions about the researcher real-time model building activity.

Since GDP can be predicted with the help of leading and coincident indicators, we have built a data set of quantitative and qualitative (survey) indicators, chosen according to their reliability and timeliness characteristics (for a similar data set see Faust *et al.*, 2005). In

Appendix 1 the variables have been divided into three subsets (I_1, I_2 and I_3) according to the length of the time series. All indicators available before 1970 belong to the first subset (I_1); the indicators with data available before 1974 (and after 1972) belong to the second subset (I_2); and the indicators with data available before 1980 (and after 1978) belong to the third subset (I_3). The data-set of the potential predictors is not organised by vintages because they are not usually revised, the only exception being the industrial production index. The Italian real-time data-set is therefore limited to the GDP at constant prices and to the index of industrial production.

For each vintage, one-quarter ahead GDP forecasts are obtained from the following alternative models (listed from simpler to more complex models): random walk (RW), ARIMA, leading-indicator (LI) and coincident-indicator (BM) bridge models. There is a growing burden of specification work from RW and the simple ARIMA to the LI and BM models. RW is a fixed-specification model based on a GDP constant-growth (drift) assumption; ARIMA models need a little more of specification search about GDP dynamics; LI and BM models are obtained from the search of the most parsimonious specification of the most suitable indicators, see Baffigi *et al.* (2004) and Golinelli and Parigi (2004). More specifically, the LI information set includes past realisations of both GDP and leading indicators only, while the BM one includes past realisations of GDP and both coincident and leading indicators.

Since the advantage of exploiting additional sources of information on GDP comes at the growing price (measured by "costly" inefficient estimates) of selecting too short a list of indicators from the I_1, I_2 and I_3 subsets above, it is crucial to find a way to "model the modeller" in the real-time search for the best specification. In this context automatic model-search procedures are an interesting option because they are replicable, are based on well defined assumptions about the steps of the search for the final model, and avoid the mixture of *ex post* knowledge (unavailable to the researcher) with purely *ex ante* data-based information (available to the researcher). In other terms, automatic model selection guards against future information creeping into the model specifications and thus into the forecasts. Two alternatives are either to use "vintage" models built in the past (see *e.g.* Fair and Shiller, 1990) or to keep model specifications fixed over time. The first one is unfeasible essentially because past models or forecasts are not available (see Busetti, 2001, for an exception in a different context), while the second one is uninteresting because we want to allow our virtual modeller to update models specification as new information is released. In this exercise we cannot replicate the researcher modelling ability (the "art" of forecasting), so that our results should be taken as a sort of "lower bound" with respect to the outcome of a proper modeller.

Of the four model typologies listed above, the RW is the only one that does not change over vintages. The ARIMA models vary by vintage and are obtained by pretesting for unit roots (see Elliott *et al.*, 1996), to choose between specifications in differences or in levels with a trend, and by using the Hannan and Rissanen (1982) procedure to find the final ARIMA models[11].

The task of finding the specification of LI and BM models for each GDP vintage is by far more challenging, but it may be greatly simplified by applying the variable-selection/model-reduction method implemented in the *PcGets* package (see Hendry and Krolzig, 2001). The *PcGets* preferred LI and BM models for each GDP vintage are assumed to be the final choice of our "virtual" modeller, and are used for one-quarter ahead forecasting exercises. *PcGets* requires some inputs about: *(a)* the list of the regressors, *(b)* the specific strategy to adopt in the reduction process from a general unrestricted model (GUM); *(c)* the more appropriate data transformations, *(d)* the lag length, and *(e)* the deterministic components.

As far as point *(a)* is concerned, in order to reduce the number of potential explanatory indicators, we split the set of GDP vintages into three blocks corresponding to the major changes in national accounts definitions (see Table 2) and to indicator data availability (see Appendix 1). Given the whole vintage availability, the first eleven and the latter two vintages are not used in modelling. We have to discard the first eleven vintages of the real-time GDP data-set because of the lack of indicators with an adequate number of observations (we assume a window of 80 quarters as the minimum sample for estimation).[12] Since the 67th vintage is the latest available, it is assumed to be the "actual" GDP, delayed one vintage with respect to the 65th vintage, the last used for models estimation.

The first block includes the GDP vintages from 12 to 30, classified in SEC 79 (1985 base-year), and modelled using all the sixteen series in I_1. The second block includes the

[11] The estimation procedure starts by fitting an 8th order autoregression to the GDP series to obtain the corresponding residuals, then all combinations of AR and/or MA components up to the 4th order are estimated (MA regressors are measured by the AR(8) residuals), and the ARIMA orders for each vintage are chosen from the combination minimising the Akaike criterion (AIC). The corresponding ARIMA models are estimated and further refined on the basis of correlogram inspection, residual autocorrelation and parameter significance tests.

[12] In this way, the first regression is estimated over the period 1971 Q2 – 1991 Q1, and is used to forecast the twelfth GDP vintage.

vintages from 31 to 50: twelve SEC 79 vintages in base 1990 and eight SEC 95, incomplete, vintages in base 1995.[13] In principle, these GDP vintages could be explained by both I_1 and I_2 information sets, but a degree-of-freedom shortage would rise. Therefore, we specify the GUM of this second block from the indicators in I_2 plus the indicators in I_1 that were retained in the final models of the last two vintages of the first block (*i.e.* the 29th and the 30th). The third block includes the SEC 95 vintages from 51 to 65 (in 1995 base-year): as above, the variables of interest in the GUM are the new indicators in the I_3 set, plus the indicators retained in the final models for the last two vintages (49th and 50th) of the second block.

As far as point *(b)* is concerned, in *PcGets* the GUM is reduced to a final parsimonious specification according to two alternative strategies: "liberal" and "conservative". The "liberal" strategy aims to keep as many as possible variables (at the risk of retaining irrelevant indicators); the "conservative" one discards irrelevant variables (at the risk of omitting important indicators). Both strategies involve a variety of tests, assumptions and choices to select the preferred model (*PcGets* applies a multi-path reduction search, after having dropped highly insignificant variables on the basis of a loose significance level). In this exercise we have adopted the "conservative" strategy (with strict significance levels; see Hoover and Perez, 1999) given that only few indicators in the I_1, I_2 and I_3 subsets should really matter.

In case of *(c)*, though alternative parameterisations of the same GUM are equivalent specifications, the algorithmic simplification from each re-parameterisation has been shown to yield different final models (see *e.g.* Campos and Ericsson, 1999). In this sense, the data transformations of a GUM may be seen as representing the modeller's value added to achieve the best and more parsimonious specifications. In our "virtual-modeller" experiments, we adopt for each vintage four alternative parameterisations, corresponding to alternative transformations of the information set: with parameterisation *A* we follow the traditional approach of taking trending variables (such as GDP and industrial production) in first differences, and non-trending variables (such as ratios, balances and rates) in levels; the parameterisation *B* corresponds to the error-correction model with levels and first differences; with parameterisation *C* all variables are transformed in first differences; in parameterisation *D* the GUM is given by the pooling of the final models of the previous three parameterisations (see Hendry and Krolzig, 2004).

[13] These eight SEC 95 vintages are incomplete because their time-series do not start later than 1970 Q1. In order to preserve the fixed sample period, initial missing data have been "backcasted" by using the growth rates of the most recent longer vintage.

Points *(d)* and *(e)* above have less problematic implications. As usual with quarterly models, we set a lag length equal to 4 for all the variables in the GUM for LI models, and equal to 3 in the GUM for BM, where the simultaneous relationships between GDP and its indicators have also been considered. Finally, each GUM always includes a constant while the time trend is included only for the GUM in levels (no dummy variables are used; all variables are seasonally adjusted).

Given that 53 vintages of GDP data are modelled with 10 alternative specifications, our whole forecasting framework is based on 530 estimated models: 53 RW models, 53 ARIMA models, 212 LI models (53 LI models for each parameterisation A, B, C and D), and 212 BM (as for LI models). The fitting performance of all the estimated models is summarised in Figure 2, where the standard errors (s_M, with M = RW, ARIMA, LI*A*, LI*B*, LI*C*, LI*D*, BM*A*, BM*B*, BM*C*, BM*D*) of the 530 regressions are reported (each line corresponds to one of the ten models), and the last observation of each rolling-sample regression is reported along the horizontal axes (the first observation being 80 quarters earlier).

Figure 2 here

In Figure 2(a) four lines are reported, one for each s_M, (M= RW, ARIMA, LI*D* and BM*D;* LI and BM models are represented only by their parameterisations *D*, whose models are valid reductions of the corresponding parameterisations *A*, *B* and *C*; see below). Since RW cannot explain anything of the GDP-growth variability, the s_{RW} declining path reflects the progressively lower GDP-growth volatility over time, and its decline is fairly homogeneous within each of the three blocks of vintages (denoted by different shaded areas). ARIMA slightly improves the RW fit, although in a declining way over time.

Figure 2(b) reports the s_{ARIMA}, $s_{\mathrm{LI}D}$ and $s_{\mathrm{BM}D}$ as a ratio over s_{RW} to simplify comparisons: the closer the lines to one (the top of the figure), the most similar (in terms of standard errors) is the performance of the model to that of the RW. The results in the table show that the "*PcGets*-automatic-modeller" makes a good job in exploiting indicators information. The marginal utility of coincident indicators to predict GDP is very evident in the second and third block of GDP vintages, while in the first block it seems that leading indicators embody enough information to improve GDP predictability over the other models.

The second line of plots in Figure 2 compares the GDP-fitting ability of *A*, *B*, *C* and *D* parameterisations within LI and BM approaches in Figures 2(c) and 2(d) respectively: with very few exceptions, parameterisations *D* (thick-solid lines) are, as expected, the best performing. Thick-dotted lines in Figures 2(c) and 2(d) show the fragility of all-in-differences models (parameterisation *C*). On the other hand, the good performance of models embodying error-correction mechanisms (parameterisation *B*) is fairly evident, while parameterisation *A* performance is mixed. We could therefore tentatively conclude that GDP fitting is often worsened by incorrectly omitting levels (as in parameterisation *C*) and that the best strategy should be to start from a GUM with alternative data transformations and let *PcGets* mix such information at a second stage.

3.2 The models ability to forecast preliminary GDP

The results of the previous paragraph are all based on in-sample evidence of predictability, which does not guarantee significant out-of-sample predictability (see Granger, 1990). In fact, the danger with the use of in-sample criteria is to detect spurious GDP predictability (*i.e.* through overfitting or data-mining), due to the application of particular specification search procedures (see Clark, 2004).

In order to assess the out-of-sample performance of our models, we conduct the following forecasting exercise. Given the RW, ARIMA, LI, and BM final specifications varying from one vintage to the other, one-quarter ahead forecasts for each vintage are obtained from a rolling procedure with a fixed estimation window (T=80). In this way, our empirical framework is adaptive in both model specification, except for the RW case, and parameter estimation[14].

The one-quarter ahead GDP growth rate predictions are compared with the first GDP outturn (growth-within) and, on the basis of the resulting one-quarter ahead forecast errors, we computed a number of alternative measures of forecast evaluation for all our models. The higher panel of Table 5 reports along the columns the mean error (ME), the mean absolute error (MAE), the root mean squared error (RMSE), and the percentage of times that the sign of the first release of the GDP growth rate is correctly predicted (PS).

Table 5 here

[14] Swanson and White (1997) find that adaptive models perform better than non-adaptive models, and Stock and Watson (1996) define rolling regressions as moderately adaptive approaches outperforming both fixed-coefficient and recursive least squares models.

We also computed a number of statistics to compare (on the basis of the RMSEs above) the forecasting ability of our models with respect to that of the RW model, considered as benchmark. The columns of the lower panel of Table 5 report: the ratio of each model RMSE on that of RW (Ratio); the p-values of both the Diebold and Mariano (1995; DM) test, adjusted following Harvey *et al.* (1997), and the Giacomini and White (2004; GW) test for the null that the RMSE of the RW is equal to that of the models listed along the rows; the percentage of times that the forecast error of each model is smaller in absolute value than that of the RW model (PL).

Given that the standard errors of our rolling regressions tend to decrease over time but are fairly homogeneous within blocks (see Figure 2), the forecasting ability statistics in Table 5 are computed for both the whole sample (1991 Q2 – 2004 Q3), and for each of the three sub-periods.

The top-down analysis of Table 5 shows that the selected short-run indicators have specific information contents for GDP: both MAE and RMSE tend to decrease from RW to BM approaches in all sub-periods (the superiority of BM forecasts is in line with the literature on bridge models; see Golinelli and Parigi, 2004). The simultaneous indicator information (already available when the GDP is forecast because they are timely released) significantly improves the performance of other models that do not exploit such additional knowledge.

It must be stressed that the out-of-sample analysis downsizes the statistical relevance of lagged information detected by the in-sample search discussed in paragraph 3.1. In fact, the forecasting ability of both ARIMA and LI models is seldom significantly better than the RW model, the opposite of what is shown in Figure 2. This result further confirms the seminal findings of Ashley *et al.* (1980) and more recently Stock and Watson (2003) and Clark (2004). In-sample Granger-causality of (leading) indicators is not enough to build models with a satisfactory forecasting performance. In fact, over the whole sample, only the LI model in levels (parameterisation *B*) does significantly better than RW; this result is valid in the first and second sub-periods, but not in the third one.

3.3 Model specification and estimation along different GDP vintages

Plots in Figure 3 summarise how LI, and BM final specifications discussed above vary from each vintage to the following. The first row of graphs reports the number of indicators selected by *PcGets* for LI and BM, and suggests a smooth decline over time of their number for LI models, while BM exploit the information of more explanatory variables in the intermediate block of vintages. In any case, our adaptive-specification approach produces remarkable model variability by vintage.

Figure 3 here

In order to deepen this point, the second row of plots in Figure 3 reports alternative realisations of an index of specification variability from one vintage to the following. The index ranges from zero (when the model specification does not change from one vintage to the previous) and one (when a specification has completely different regressors with respect to the previous vintage model).[15] Data revisions embodied in each new GDP vintage induce, on average, a change of about 50% of the regressors of the previous-vintage specification, LI models being more affected than BM. The specification is unchanged from one vintage to the following (*i.e.* the index is zero) only in 8 out of 212 cases for LI models (in 39 cases for BM), while the maximum-variability level (*i.e.* the index is one) is reached 18 times for LI models, never for BM.

Regarding the usefulness (measured by in-sample fitting) of such specification variability, the plots a, b and c of Figure 4 report the ratios (minus one) between the standard error of specifications varying from one vintage to the other (at the numerator), and the standard error of regressions where the previous-vintage specification - either one-quarter or one-year earlier - is estimated using the most recent data (at the denominator). In general, we note that the in-sample fitting gain from updating model specification along vintages (*i.e.* negative ratios in Figure 4) is bigger for BM than for ARIMA models. The bigger gains are clustered in the first block of vintages for LI models and in the second block for BM, *i.e.* when the number of indicators is larger. Annual specification updates imply costs that are slightly more persistent over time than in the quarterly specification updates.

Figure 4 here

[15] In particular, the index of specification variability (from vintage *v-1* to *v*) is a ratio. The numerator is the sum of the number of regressors of the model for vintage *v-1* that are no longer in model *v*, plus the number of new regressors of the model for vintage *v* that were not in *v-1*. The denominator is the sum of the number of regressors in models *v* and *v-1*. If the specification from one vintage to the following changes, the numerator of the ratio (and then the index) is zero; if the specification for *v* has nothing in common with that for *v-1*, numerator equals denominator (the ratio is one).

Although the prevalence of negative ratios in Figures 4a-4c (their averages over time are always negative) may be interpreted as a suggestion to update both model specification and parameter estimation as soon as a new data vintage is available, the existence of positive ratios and, mainly, the purely in-sample nature of such ratios suggest to deepen the analysis by performing a number of out-of-sample forecasting exercises. In particular, Figure 4d compares the RMSE of 1-quarter ahead forecasting exercises obtained by updating Lid and BMd specification and estimation for each vintage[16] (labelled “1” in histogram 4d), with the RMSEs from alternative lags in model specification-estimation updating (labelled “2-5” in the histogram 4d).

Going from case “1” to case “5” lags in updating models are growing: cases “2” and “3” only update parameter estimates of old (1-quarter and 1-year earlier) specifications, while cases “4” and “5” use new issued data without changing neither the previous (old) specification nor parameter estimates. As expected, histogram 4d suggests growing RMSEs from progressively ignore new vintage information, and confirms the better quality (in terms of forecasting ability) of BM over LI, independently of the amount of the updating activity. However, the small difference between “1” and “2, 3” RMSEs reduces the emphasis on changing specification over vintages, while relatively large “4, 5” RMSEs stress that, independently of the specification used, model parameters have to be re-estimated , as soon as new data are available.

These RMSE outcomes can be explained by a mixture of econometric-statistical facts and of institutional facts peculiar to the Italian database. On one hand, the indicators we use are quite collinear each other, and our models are able to capture mainly short-run relationships between GDP and indicators: all these facts suggest bigger risks of breaks in parameter estimates instead of problems in finding the better specification. On the other hand, it is worth to remember that the Italian statistical agency usually revises GDP data as new information is made available, while almost all Italian indicators are not revised: since both LI and BM models entail relationships between GDP (changing over vintages) and indicators (mainly fixed), parameter changes are quite likely all times new GDP data are released.

4. Outturns vs. models to forecast “actual” GDP

Analysts usually compare the preliminary estimates of GDP with the most recent one-step ahead forecasts, obtained on the basis of indicators and the evolution of GDP according

[16] These are the same exercises we made and reported in the previous section.

to the previous vintage. Although it is known that these estimates will be revised with the next vintage, we are not aware of attempts to represent the evolution of actual GDP, that is the evolution which should not be revised further. This is basically due to the fact that there does not exist a series called actual GDP. Statistics always vary, because of change in definitions, improvements in statistical practice and so on. However, by taking into account the average forecasting horizon (3 to 5 years), we can try to define a sort of "actual" GDP and thereby assess the performance of real-time GDP forecasts with respect to this series.

"Actual" GDP may be defined in two alternative ways. The most obvious one is to define actual GDP as the latest available vintage (in our case, the 67th vintage ending in 2004 Q4). The difference between this "actual" GDP and the first outturn, or the "forecast error" of the first outturn (FOE), embodies both statistical and benchmark revisions occurred in the period between the two releases. Clearly for old vintages the weight of benchmark revisions is higher, while it tends to decline for younger ones.

In order to disentangle the portion of the FOE due only to statistical changes (due to new information availability), a second definition of actual GDP is proposed: the intermediate outturn, or the last vintage before a benchmark revision occurs. According to the three blocks of vintages described in Section 3, we have three different actual GDP, defined as intermediate outturn: the 30th vintage for the first block (from 1970 Q1 to 1995 Q3), the 50th vintage for the second (from 1995 Q4 to 2000 Q3), and the 67th vintage for the third (from 2000 Q4 to 2004 Q4). In this last case, the latest vintage and the intermediate outturn coincide and the FOE represents only statistical changes.

Table 6 summarises main statistics about the ability of our models and of the first outturn in forecasting the intermediate GDP outturn, or the different performances with respect to statistical changes.

Table 6 here

The first outturn appears to perform always better than the models based only on past information (LI*A*-*D*). The forecast ability of the various BM models is instead similar to that of the first outturn (lower panel of Table 6); this result is even more evident when we consider the three sub-samples of our exercise. More specifically, the BM forecasting performance is particularly good in the first period and slightly worsens in the following two; the percentage of BM forecast errors smaller than the FOE (PL in Table 6) declines, on average, from one-half in the first period to one-third in the third period. This evidence may be interpreted as a sign of the growing ability of national accounts to embody relevant statistical information in the very short-run, despite the increased timeliness of GDP releases.

Table 7 reports the results on the forecasting ability of our models, the first and the intermediate outturns in predicting the latest available GDP vintage, which not only accounts for statistical changes but also embodies definitional changes.

Table 7 here

As in Table 6 the first outturn performance appears to be better than that of LI models (although not so clearly); however, the statistics of the FOE are fairly in line with those of the BM. In general, the first outturn forecasting accuracy is only occasionally significantly better than all the model-based predictions. Finally and surprisingly, the intermediate outturn (that embodies all short-run statistical information) forecasting performance appears to be better than all other predictions only in the second sub-sample.

From the evidence presented in Tables 6 and 7 we can tentatively conclude that the statistical changes are well predicted by the first outturn and by the models based on coincident information (BM). However, if we also consider definitional changes (i.e. we consider the latest available vintage), the GDP path becomes less predictable. The RMSE of the first outturn worsens and approaches that of the BM models (their RMSE ratio is much closer to one).

In cases where two competing models appear to be characterised by a good forecasting performance, it may be interesting to assess whether the predictions from one model may be encompassed by the other. In other terms, we want to assess whether the informative content of the predictions of one model already embody that of the competitor (which can therefore be ignored; see *e.g.* Granger and Newbold, 1986). In our case, this implies another sort of comparison between the first outturn and the BM models based on the forecast encompassing test proposed by Fair and Shiller (1990). The test can be conducted with reference to the latest available case or the intermediate outturn case by running the following regression:

$$(1) \qquad \Delta\log(AC\_GDP_t) = \alpha + \beta\,\Delta\log(FIRST\_GDP_t) + \gamma\,\Delta\log(MODEL\_GDP_t) + u_t$$

where: Δ is the first-difference operator; AC_GDP_t is either the latest available or the intermediate GDP vintage; $FIRST\_GDP_t$ is the first outturn; $MODEL\_GDP_t$ is the one-quarter ahead GDP forecast obtained by each of the described models in paragraph 3.1; u_t is the error term, which, as noted in Fair and Shiller (1990), may be possibly autocorrelated and/or heteroskedastic. The regression is computed over the whole sample (from 1991 Q2 to 2004 Q3) because inferences in the three sub-samples would be affected by an excessive degrees of freedom shortage.

A number of features make equation (1) particularly suitable for our interests: it is in first differences (accounts for both the GDP log-levels non-stationarity and the results in paragraph 2.3); the β and γ parameters are not constrained to sum to one (see Fair and Shiller, 1990); the α parameter allows for bias-correction (biased forecasts can be evaluated); regressors and regressand are the same as those used to compute the statistics in Tables 5 to 7 (no further data transformations are required).

We are mainly interested in the following cases (from the less to the most favourable to the BM models): *(a)* $\beta \neq 0$ and $\gamma = 0$, the information of model-based forecasts is completely ecompassed by the first outturn; *(b)* $\beta \neq 0$ and $\gamma \neq 0$, both model-based forecasts and the first outturn contain independent relevant information for the prediction of the dependent variable; *(c)* $\beta = 0$ and $\gamma \neq 0$, the information of the first outturn is encompassed by the model-based forecasts.

The Fair-Shiller test-statistics for all these cases is the t-statistics of the OLS estimated parameter of equation (1). Table 8 reports the t-statistics of β and γ OLS estimates by using, along the columns, three alternative estimators of the variance-covariance matrix of the residuals: the identity matrix (equivalent to the assumption of i.i.d. errors) in the second and third columns; the White (1980) heteroskedasticity-consistent estimator in the fourth and fifth columns; the Newey and West (1987) heteroskedasticity- and autocorrelation-consistent estimator in the last two columns.

Table 8 here

In all cases the Fair and Shiller test results are unambiguous and robust to the choice of the variance-covariance estimator. The first outturn encompasses the forecasts from ARIMA and LI models (only β estimates are always significant). However, when the first outturn is compared with the forecasts from BM models, we have mixed results: β parameter estimates are largely significant in the regressions where AC_GDP is the intermediate-outturn vintage, while their significance tends to disappear when the AC_GDP is the latest available vintage; the opposite occurs for the significance of the γ estimates. It appears therefore that both the first outturn and the BM forecasts carry useful information for actual GDP growth, the first outturn being more relevant for the prediction of the internmediate-outturn, and the BM forecasts for the prediction of the latest available vintage. In other words, practitioners should not discard their BM-based forecasts as soon as preliminary GDP data are issued; rather, our results suggest that a more complete assessment of the actual economic situation can be achieved by a combination of the two sets of data.

5. Conclusions

As soon as new GDP data are released by statistical agencies, practitioners: *(a)* compute the GDP growth rate (disregarding its levels), and *(b)* consider that growth rate as the best GDP picture available for that quarter, at least until a second outturn (available one-quarter later) revises preliminary figures. In this paper we have tried to assess the statistical foundations of practices *(a)* and *(b)* above, by using a real-time quarterly data set for Italian GDP from the vintage 1970 Q1 - 1988 Q2 to the latest available one 1970 Q1 – 2004 Q4.

As far as point *(a)* is concerned, the levels/growth rates trade-off is solved in favour of the second one with the outcomes of the cointegration analysis (the levels of interest are generated by unit-root processes). However, an appropriate GDP measure for levels is not available, as data published over time are heterogeneous and are characterised by benchmark revisions (such as base and classification changes). In addition, the concept of "finally revised" GDP data is not uniquely determined.

Regarding the construction of consistent GDP levels, we suggest two alternative methods. With the first method, GDP levels in each released time series are rescaled for the initial value, delivering a sort of index 1970 Q1-based (even in this case the growth rates of the various GDP outturns are not coherent with the officially published GDP growth rates, because of residual heterogeneity). With the second method, GDP levels are obtained by cumulating GDP growth rates as they are issued starting from the first outturn (when data are published for the first time); the initial values of such procedure are supposed to coincide with the first vintage of the rescaled data (from previous method).

As far as the issue of measuring the finally revised GDP is concerned, we opted for two definitions: the latest vintage published before the occurrence of a benchmark change, *i.e.* a sort of intermediate-outturn; the latest available vintage. The differences between the preliminary release and the intermediate-outturn reflect only statistical changes, while the differences between the preliminary release and the latest available data account for all sources of change, both statistical (due to new available information) and definitional (due to new base-year and/or new accounting schemes).

Several cointegration exercises are robust in suggesting that GDP levels (however computed) are not needed to characterise the Italian GDP revision process. Moreover, it emerges that latest available GDP levels do not share any long run trends with GDP levels issued during the first two years. Therefore, probably due to a lack of satisfactory ways to define GDP levels bearing useful information across vintages, practitioners are right in computing growth rates and in ignoring the corresponding levels.

The assessment of the quality of the first outturn in predicting the corresponding actual figures (point *(b)* above; “actual” means either the latest available or the intermediate outturn) requires the construction of forecasting models acting as antagonists of the statistical agency first preliminary estimates. 10 alternative models have been considered, from univariate to multivariate specifications, with both leading and coincident indicators, generating a total of 530 one-quarter ahead model-based forecasts. An original characteristic of our analysis is that it is based on a sort of real-time exercise, where we try to reproduce the usual forecasting practice when new data are available by comparing previous forecast with the official data and by updating models in order to take account of the new pieces of information. This has required not only to model the data, but the forecaster behaviour as well. The main danger in this kind of exercise is that it is difficult to limit the influence during the specification process of the future evolution of a variable when this is already known. The availability of automatic model building procedures, such as *PcGets*, avoids this problem, although it cannot reproduce the ability, or the art, of an actual forecaster.

A first set of exercises suggest that bridge models using both coincident and leading indicators significantly outperform alternative models in forecasting GDP first release growth rates, thus confirming in the real-time case a common result in the more traditional bridge models literature (see Golinelli and Parigi, 2004).

In a second set of experiments we compare the ability of our ten models with that of the first GDP release in forecasting the actual growth rates. The most striking result is that bridge models forecasts appear to embody information that is not completely accounted for by the first GDP release, especially when actual is supposed to mean the latest available growth rates. This result is similar to that obtained by Faust *et al.* (2005), when they show that latest available figures are related to preliminary ones and to some indicators. However, our results differ in an important qualitative feature. While Faust *et al.* apply a sort of *ex-post* analysis, which is informative *per se* but has no direct impact on the real-time work of forecasters, our *ex-ante* exercises have potentially relevant implications for the researchers, in particular when initial conditions of a forecasting exercise have to be established. In this case, our results suggest that a better description of the actual evolution of GDP can be achieved by combining the most recent one-quarter ahead forecasts with the corresponding preliminary figures (while our forecast come from bridge models, that is from a linear combination of short-term indicators, Busetti, 2001, provides some evidence for the combination of the preliminary estimates with the forecast from a large macroeconometric model).

References

Ashley, C., C. W. J. Granger and R. Schmalensee (1980), "Advertising and aggregate consumption: an analysis of causality", *Econometrica*, Vol. 48, No 5, pp. 1149-1168.

Baffigi, A., R. Golinelli and G. Parigi (2004), "Bridge Models to Forecast the Euro Area GDP", *International Journal of Forecasting*, Vol. 20, No 3, pp. 447-460.

Bordignon, S. and U. Trivellato (1989), "The optimal use of provisional data in forecasting with dynamic models", *Journal of Business & Economic Statistics*, Vol. 7, No 2, pp. 275-286.

Busetti, F. (2001), "The use of preliminary data in econometric forecasting: an application to the Bank of Italy quarterly model", Bank of Italy, *Temi di Discussione*, No 437.

Campos, J. and N. R. Ericsoon (1999), "Constructive data mining: modelling consumers' expenditure in Venezuela", *Econometrics Journal*, Vol. 2, pp. 226-240.

Castle, J. and C. Ellis (2002), "Building a real-time database for GDP(E)", Bank of England, *Quarterly Bulletin*, Spring, pp. 42-49.

Clark, T. E. (2004), "Can out-of-sample forecast comparisons help prevent overfitting?", *Journal of Forecasting*, Vol. 23, pp. 115-139.

Clausen, J. R. and C.-P. Meier (2003), "Did the Bundesbank follow a Taylor rule? An analysis based on real-time data", University of Cologne, *IWP Discussion Paper*, No 2, August.

Croushore, D. and T. Stark (2003), "A real-time data set for macroeconomists: does the data vintage matter?", *The Review of Economics and Statistics*, Vol. 85, No 3, pp. 605-617.

de la Rocque Palis, R., R. L. O. Ramos and P. Robitaille (2004), "News or noise? An analysis of Brazilian GDP", The Federal Reserve Board, *International Finance Discussion Papers*, No 776, revised.

Diebold, F. X. and R. S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-63.

Di Fonzo T., S. Pisani and G. Savio (2002), "Revisions to Italian quarterly national accounts aggregates: some empirical results", *INSEE-Eurostat Workshop on Quarterly National Accounts*, Luxembourg, Eurostat, pp. 7-31.

Dopke, J. (2004), "Real-time data and business cycle analysis in Germany", Deutsche Bunsedbank, *Discussion Paper*, Series 1, No 11.

Elliott, G., T., J. Rothemberg, and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, Vol. 64, No 4, pp. 813-836.

Engle, R. F. e C. W. J. Granger (1987), "Cointegration and error correction: representation, estimation and testing", *Econometrica*, Vol. 55, pp. 251-276.

Fair, R. C. and R. J. Shiller (1990), "Comparing information in forecasts from econometric models", *The American Economic Review*, Vol. 80, No 3, pp. 375-389.

Faust, J., J. H. Rogers and J. H. Wright (2005), "News and noise in G-7 announcements", *Journal of Money, Credit and Banking*, forthcoming

Gallo, G. M. and M. Marcellino (1999), "Ex post and ex ante analysis of provisional data", *Journal of Forecasting*, Vol. 19, pp. 421-433.

Garratt, A. and S. P. Vahey (2004), "UK real-time macro data characteristics", *Birkbeck Working Papers in Economics & Finance*, No 0513.

Giacomini, R. and H. White (2003), "Tests of Conditional Predictive Ability", University of California, San Diego, *Discussion Papers*, No 2003-09.

Giovannini, E. (1993), "L'informazione statistica per le decisioni: alcune linee generali e un esempio illustrativo", ISTAT, *Quaderni di Ricerca*, Interventi e Relazioni.

Golinelli, R. and G. Parigi (2004), "Using Monthly Indicators to Bridge-Forecast Quarterly GDP for the G7 Countries", downloadable at the URL http://www.dse.unibo.it/golinelli/research/GoPa_G7_EN_4.pdf

Granger, C. W. J. (1990), *Modelling Economic Series. Readings in Econometric Methodology*, Oxford University Press.

Granger, C. W. J. and P. Newbold (1986), *Forecasting Economic Time Series* (2^{nd} edition), Academic Press.

Hannan, E. J. and J. Rissanen (1982), "Recursive estimation of mixed autoregressive-moving average order", *Biometrika*, Vol. 69, No 1, pp. 81-94.

Harvey, D. I., S. J. Leybourne and P. Newbold (1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, Vol. 13, pp. 281-91.

Hendry, D. F. and H.-M. Krolzig (2001), *Automatic Econometric Model Selection Using PcGets 1.0*, Timberlake Consultants Press.

Hendry, D. F. and H.-M. Krolzig (2004), "Resolving three 'intractable' problems using Gets approach", Oxford University, August, *mimeo*.

Hoover, K. D. and S. J. Perez (1999), "Data mining reconsidered: encompassing and the general-to-specific approach to specification search", *Econometrics Journal*, Vol. 2, pp. 167-191.

Koenig, E. F., S. Dolmas, and J. Piger (2003), "The use and abuse of real-time data in economic forecasting", *The Review of Economics and Statistics*, Vol. 85, No 3, pp. 618-628.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Lupi, C. and F. Peracchi (2003), "The limits of statistical information: how important are GDP revisions in Italy?", in F. Kostoris (ed.), *Monitoring Italy*, forthcoming.

Mankiw, N. G, D. E. Runkle, and M. D. Shapiro (1984), "Are preliminary announcements of the money stock rational forecasts?", *Journal of Monetary Economics*, Vol. 14, pp. 15-27.

Mankiw, N. G, and M. D. Shapiro (1986), "News or noise? An analysis of GNP revisions", *NBER Working Paper*, No 1939, June.

Mork, K. A. (1987), "Ain't behavin': forecast errors and measurement errors in early GNP estimates", *Journal of Business & Economic Statistics*, Vol. 5, No 2, pp. 165-175.

Newey, W. and K. West (1987), "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, Vol. 55, pp. 703-708.

Pain, N. (1994), "Cointegration and forecast evaluation: some lessons from National Institute forecasts", *Journal of Forecasting*, Vol. 13, pp. 481-494.

Patterson, K. D. (2000), "Which vintage of data to use when there are multiple vintages of data? Cointegration, weak exogeneity and common factors", *Economics Letters*, Vol. 69, pp. 115-121.

Patterson, K. D. (2002), "The data measurement process for UK GNP: stochastic trends, long memory, and unit roots", *Journal of Forecasting*, Vol. 21, No 4, pp. 245-264.

Patterson, K. D. (2003), "Exploiting information in vintages of time-series data", *International Journal of Forecasting*, Vol. 19, pp. 177-197.

Patterson, K. D. and S. M. Haravi (1991), "Data revisions and the expenditure components of GDP", *The Economic Journal*, Vol. 101, No 407, pp. 887-901.

Rettore, E. and U. Trivellato (1986), "Preliminary data errors and their impact on the forecast error of simultaneous equations models", *Journal of Business & Economic Statistics*, Vol. 4, No 4, pp. 445-453.

Roodenburg, O. (2004), "On the predictability of GDP data revisions in the Netherlands", De Nederlandsche Bank, *Working Paper*, No 4, July.

Siklos, P. L. (1996), "An empirical exploration of revisions in US national income aggregates", *Applied Financial Economics*, Vol. 6, pp. 59-70.

Steckler, H. O. (1967), "Data revisions and economic forecasting", *Journal of the American Statistical Association*, Vol. 62, No 318, pp. 470-483.

Stock, J. H. and M. W. Watson (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations", *Journal of Business Economics and Statistics*, Vol. 14, p. 11-30.

Stock, J. H. and M. W. Watson (2003), "Forecasting Output and Inflation: the Role of Asset Prices", *Journal of Economic Literature*, Vol. XLI, pp. 788-829.

Swanson, N. R., and D. van Dijk (2004), "Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry", Erasmus University (Rotterdam), *Econometric Institute Report*, No 28 (2001), revised 2004.

Swanson, N. R., and H. White (1997), "A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks", *The Review of Economics and Statistics*, Vol. 79, No 4, pp. 540-550.

White, H. (1980), "A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity", *Econometrica*, Vol. 48, pp. 817-838.

Appendix 1: The indicators database

THE THREE SUBSETS OF INDICATORS: I_1, I_2 AND I_3

The first subset of indicators (I_1): first observation available before 1970

xp^v	vintages of industrial production index (logs of index levels) (1) (3)
glav	working days (log-levels)
dpe	stock prices growth rate (first-differences of log-prices)
dpoil	oil price growth rate (first-differences of log-prices)
r3m	nominal 3-months interest rate (levels)
q01	rate of capacity utilisation (log-levels)
x_3	orders level: total manufacturing (2)
x_{118}	passenger cars registered (log-levels)
x_{13}	finished goods stocks: total manufacturing (2)
x_{18}	production actual tendency: intermediate goods (2)
x_{19}	production actual tendency: investment goods (2)
x_{20}	production actual tendency: consumption goods (2)
x_{21}	orders tendency (3-4 months expected demand): total manufacturing (2)
x_{26}	production future tendency (3-4 months): intermediate goods (1) (2)
x_{27}	production future tendency (3-4 months): investment goods (1) (2)
x_{28}	production future tendency (3-4 months): consumption goods (1) (2)

The second subset of indicators (I_2): first observation available before 1974 and after 1972

x_6	orders level: intermediate goods (2)
x_9	orders level: investment goods (2)
x_{12}	orders level: consumption goods (2)
x_{14}	finished goods stocks: intermediate goods (2)
x_{15}	finished goods stocks: investment goods (2)
x_{16}	finished goods stocks: consumption goods (2)
x_{22}	orders inflow / demand tendency (3-4 months):intermediate goods (1) (2)
x_{23}	orders inflow / demand tendency (3-4 months): investment goods (1) (2)
x_{24}	orders inflow / demand tendency (3-4 months): consumption goods (1) (2)
q10	consumer confidence (logs of index levels)

The third subset of indicators (I_3): first observation available before 1980 and after 1978

x_{34}	future tendency of the economy, 3-4 months (Isae): intermediate goods (2)
x_{35}	future tendency of the economy, 3-4 months (Isae): investment goods (2)
x_{36}	future tendency of the economy, 3-4 months (Isae): consumption goods (2)
x_{55}	production actual tendency (Isae): investment goods (1) (2)
x_{61}	production actual tendency (Isae): consumption goods (1) (2)
x_{67}	production actual tendency (Isae): intermediate goods (1) (2)
x_{50}	orders actual tendency (Isae): investment goods (1) (2)
x_{56}	orders actual tendency (Isae): consumption goods (1) (2)
x_{62}	orders actual tendency (Isae): intermediate goods (1) (2)
q11	manufacturing and construction confidence, weighted average (logs of index levels)
reale	real interest rate on bank loans (levels)

(1) Seasonally adjusted data. (2) $x_i = log(1+IM_i /100)$, where IM_i are quarterly averages of monthly survey data, i is the code reported in the first column. (3) Only the industrial production database is organised in vintages: in Italy, survey and financial data are not subject to revisions.

Table 1

THE REAL-TIME DATA SET FOR GDP LEVELS, ${}^{o}y^{v}_{t}$ [1]

period, t \ vintage, v	1	2	3	4		f
1	${}^{T+1}y^{1}_{1}$	${}^{T+2}y^{2}_{1}$	${}^{T+3}y^{3}_{1}$	${}^{T+4}y^{4}_{1}$	....	${}^{T+f}y^{f}_{1}$
2	${}^{T}y^{1}_{2}$	${}^{T+1}y^{2}_{2}$	${}^{T+2}y^{3}_{2}$	${}^{T+3}y^{4}_{2}$	....	${}^{T+f-1}y^{f}_{2}$
3	${}^{T-1}y^{1}_{3}$	${}^{T}y^{2}_{3}$	${}^{T+1}y^{3}_{3}$	${}^{T+2}y^{4}_{3}$	....	${}^{T+f-2}y^{f}_{3}$
4	${}^{T-2}y^{1}_{4}$	${}^{T-1}y^{2}_{4}$	${}^{T}y^{3}_{4}$	${}^{T+1}y^{4}_{4}$	....	${}^{T+f-3}y^{f}_{4}$
....	....	....	....	....	....	....
T-1	${}^{3}y^{1}_{T-1}$	${}^{4}y^{2}_{T-1}$	${}^{5}y^{3}_{T-1}$	${}^{6}y^{4}_{T-1}$	....	${}^{f+2}y^{f}_{T-1}$
T	${}^{2}y^{1}_{T}$	${}^{3}y^{2}_{T}$	${}^{4}y^{3}_{T}$	${}^{5}y^{4}_{T}$	....	${}^{f+1}y^{f}_{T}$
T+1	${}^{1}y^{1}_{T+1}$	${}^{2}y^{2}_{T+1}$	${}^{3}y^{3}_{T+1}$	${}^{4}y^{4}_{T+1}$	....	${}^{f}y^{f}_{T+1}$
T+2	n.a.	${}^{1}y^{2}_{T+2}$	${}^{2}y^{3}_{T+2}$	${}^{3}y^{4}_{T+2}$	....	${}^{f-1}y^{f}_{T+2}$
T+3	n.a.	n.a.	${}^{1}y^{3}_{T+3}$	${}^{2}y^{4}_{T+3}$	....	${}^{f-2}y^{f}_{T+3}$
T+4	n.a.	n.a.	n.a.	${}^{1}y^{4}_{T+4}$	....	${}^{f-3}y^{f}_{T+4}$
....	....	....	....	....	....	....
T+f	n.a.	n.a.	n.a.	n.a.	....	${}^{1}y^{f}_{T+f}$

([1]) where o = 1, 2, ..., T+f is the outturn index, v = 1, 2, ..., f is the vintage index, and t = 1, 2, ..., T+f is the period index; f labels the final vintage. *n.a.* means not available (no data released for that period by that vintage).

Table 2

THE DEFINITION OF THE BENCHMARK VINTAGES

Blocks of vintages [1]	Classification	Measurement	Base	Start date [2]	Trading days
from 1 to 11	SEC 79	billion lire	1980	1970 Q1	not adjusted
from 12 to 30	SEC 79	billion lire	1985	1970 Q1	not adjusted
from 31 to 42	SEC 79	billion lire	1990	1970 Q1	not adjusted
from 43 to 50	SEC 95	billion lire	1995	1982 Q1	not adjusted
from 51 to 54	SEC 95	billion lire	1995	1970 Q1	not adjusted
from 55 to 59	SEC 95	million euro	1995	1970 Q1	not adjusted
from 60 to 67	SEC 95	million euro	1995	1980 Q1	adjusted

([1]) Each block includes all the GDP vintages between two consecutive benchmark revisions. ([2]) All the times that the period 1970 Q1 is not the start date of a vintage, missing data are "backcasted" by using the growth rates of the most current available longer vintage.

Table 3

UNIT ROOT TESTS FOR SELECTED REVISIONS
1989 Q2 – 2003 Q1

		DF-GLS [1]	lags	DF-GLS [1]	lags
		Sequential revisions			
o	o′	$\log({}^{o}x / {}^{o'}x)$ [2]		$\log({}^{o}z / {}^{o'}z)$ [2]	
1	2	-2.393 *	3	-1.729	0
2	3	-0.862	3	-0.949	4
3	4	-2.255 *	3	-0.280	0
4	5	-7.912 **	0	-1.494	1
5	6	-3.475 **	4	-1.542	0
6	7	-2.070 *	3	-0.920	1
7	8	-2.781 **	3	-0.758	0
1	4	-1.407	3	-1.788	1
4	8	-2.967 **	0	-1.831	0
		Total revisions			
o	v=f	$\log({}^{o}x / x^{v})$ [2]		$\log({}^{o}z / z^{v})$ [2]	
1	67	-0.914	2	-0.718	2
2	67	-1.334	0	-0.698	2
3	67	-0.986	2	-1.046	2
4	67	-1.064	2	-0.462	3
5	67	-0.938	3	-0.110	2
6	67	-1.848	0	-0.594	2
7	67	-1.728	0	-0.921	3
8	67	-1.422	1	-0.377	3

([1]) DF-GLS is the Elliott, Rothemberg and Stock (1996) test statistic with AIC lag-selection from max-lags=5. Critical values (p-value): -2.607 (1%), -1.947 (5%), -1.613 (10%). * and ** reject the unit root null at the 5% and 1% respectively. ([2]) x- and z- outturns are defined in section 2.1.

Table 4

JOHANSEN COINTEGRATION FOR SUB-SETS OF VARIABLES
1989 Q2 – 2003 Q1

Variables in VAR [1]:	# of lags	rank [2] r=	long run	weak exogeneity for the variable(s):	overall p-values [3]
from ${}^{1}x$ to ${}^{8}x$	3	7 *	pairwise 1, -1	${}^{8}x$	0.128 [14]
${}^{1}x$, ${}^{4}x$, ${}^{8}x$	3	2 *	pairwise 1, -1	${}^{8}x$	0.064 [4]
${}^{1}x$, ${}^{8}x$, x^{67}	4	1*	1, -1, 0	x^{67}	0.511 [4]
${}^{1}z$, ${}^{4}z$, ${}^{8}z$	3	0	-	-	-
${}^{1}z$, ${}^{8}z$, x^{67}	3	0	-	-	-
${}^{1}x$, ${}^{8}x$, ${}^{1}z$, ${}^{8}z$, x^{67}	5	1**	1, -1, 0, 0, 0	${}^{8}x$, ${}^{1}z$, ${}^{8}z$, x^{67}	0.174 [8]

([1]) All the variables listed below are in logs; x- and z- outturns are defined in section 2.1. ([2]) ** and * indicate rejection at 1% and 5% of the null that the cointegration rank is at most equal to r-1. ([3]) P-value of both the long run and the weak exogeneity over-identifying restrictions (whose corresponding number is in squared brackets).

Table 5

ONE-QUARTER AHEAD FORECASTING ABILITY OF REAL-TIME (1ST OUTTURN) GDP [1]

	Overall (91Q2-04Q3)				1st block (91Q2-95Q4)				2nd block (96Q1-00Q4)				3rd block (01Q1-04Q3)			
Forecasting models	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS
Random walk model (RW)	-0.233	0.483	0.620	74.1	-0.331	0.723	0.840	73.7	-0.123	0.388	0.514	75.0	-0.256	0.304	0.373	73.3
ARIMA model	-0.164	0.520	0.655	74.1	-0.237	0.760	0.878	73.7	-0.001	0.414	0.526	80.0	-0.288	0.358	0.447	73.3
Leading indicators A (LIa)	-0.210	0.504	0.656	77.8	-0.337	0.762	0.893	73.7	-0.182	0.446	0.579	80.0	-0.088	0.254	0.298	80.0
Leading indicators B (LIb)	-0.045	0.413	0.530	79.6	0.021	0.533	0.693	84.2	-0.004	0.335	0.411	75.0	-0.183	0.365	0.419	80.0
Leading indicators C (LIc)	0.010	0.467	0.632	85.2	-0.004	0.690	0.889	84.2	0.070	0.416	0.479	85.0	-0.053	0.252	0.363	86.7
Leading indicators D (LId)	-0.050	0.418	0.547	81.5	0.020	0.588	0.730	84.2	-0.063	0.361	0.467	80.0	-0.120	0.279	0.333	80.0
Bridge model A (BMa)	-0.163	0.328	0.434	87.0	-0.255	0.476	0.592	79.0	-0.154	0.259	0.351	95.0	-0.058	0.231	0.266	86.7
Bridge model B (BMb)	-0.105	0.324	0.397	87.0	-0.215	0.427	0.500	89.5	0.008	0.290	0.357	95.0	-0.114	0.240	0.286	73.3
Bridge model C (BMc)	-0.092	0.331	0.446	92.6	-0.342	0.479	0.628	84.2	0.119	0.247	0.284	100.0	-0.056	0.254	0.332	93.3
Bridge model D (BMd)	-0.118	0.318	0.404	90.7	-0.223	0.433	0.513	89.5	-0.073	0.283	0.384	95.0	-0.044	0.219	0.238	86.7
Comparison with RW [2]	Ratio	DM	GW	PL	Ratio	DM	GW	PL	Ratio	DM	GW	PL	Ratio	DM	GW	PL
ARIMA model	1.056	0.399	0.400	42.6	1.045	0.680	0.609	42.1	1.024	0.816	0.858	45.0	1.200	0.039	0.106	40.0
Leading indicators A (LIa)	1.057	0.469	0.565	53.7	1.063	0.550	0.598	63.2	1.127	0.588	0.573	35.0	0.800	0.145	0.204	66.7
Leading indicators B (LIb)	0.854	0.044	0.125	53.7	0.825	0.072	0.205	52.6	0.800	0.034	0.113	60.0	1.124	0.479	0.490	46.7
Leading indicators C (LIc)	1.019	0.835	0.880	51.9	1.058	0.643	0.743	57.9	0.931	0.491	0.618	35.0	0.973	0.944	0.927	66.7
Leading indicators D (LId)	0.881	0.232	0.313	63.0	0.868	0.345	0.384	68.4	0.909	0.67	0.714	60.0	0.893	0.476	0.557	60.0
Bridge model A (BMa)	0.699	0.001	0.000	61.1	0.704	0.002	0.006	73.7	0.683	0.011	0.049	55.0	0.713	0.125	0.177	53.3
Bridge model B (BMb)	0.640	0.005	0.002	64.8	0.595	0.021	0.018	78.9	0.695	0.055	0.070	60.0	0.767	0.194	0.295	53.3
Bridge model C (BMc)	0.719	0.002	0.008	63.0	0.747	0.016	0.059	68.4	0.552	0.007	0.042	50.0	0.891	0.564	0.665	73.3
Bridge model D (BMd)	0.650	0.004	0.002	61.1	0.610	0.029	0.021	68.4	0.747	0.035	0.069	60.0	0.639	0.049	0.079	53.3

(1) Statistics of forecasting ability (in %): ME (mean error), MAE (mean absolute error), RMSE (root mean squared error), PS (% of times the sign of the 1st outturn GDP growth rate is rightly predicted). (2) The forecasting ability comparison with respect to the benchmark (RW model) is RMSE-based: ratio between RMSEs (Ratio); p-values of both the Diebold and Mariano (1995) test (DM) and the Giacomini and White (2003) test (GW) for the null that the RMSE of the model is equal to that of benchmark RW; and % of times the absolute error of the model is lower than the benchmark RW (PL).

Table 6

FORECASTING ABILITY OF THE GDP INTERMEDIATE OUTTURN [1]

	Overall (91Q2-04Q3)				1st block (91Q2-95Q4)				2nd block (96Q1-00Q4)				3rd block (01Q1-04Q3)			
Forecasting models	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS
Random walk model (RW)	-0.198	0.496	0.654	74.1	-0.274	0.720	0.863	73.7	-0.068	0.401	0.573	75.0	-0.216	0.336	0.432	73.3
ARIMA model	-0.129	0.511	0.680	77.8	-0.181	0.699	0.889	78.9	0.055	0.426	0.579	80.0	-0.248	0.356	0.483	73.3
Leading indicators A (LIa)	-0.176	0.528	0.700	77.8	-0.281	0.810	0.978	73.7	-0.127	0.457	0.581	80.0	-0.048	0.379	0.450	80.0
Leading indicators B (LIb)	-0.010	0.462	0.600	81.5	0.078	0.599	0.780	89.5	0.051	0.414	0.500	75.0	-0.143	0.450	0.555	80.0
Leading indicators C (LIc)	0.044	0.514	0.685	85.2	0.053	0.779	0.946	84.2	0.125	0.428	0.548	85.0	-0.014	0.346	0.476	80.0
Leading indicators D (LId)	-0.015	0.511	0.645	81.5	0.076	0.756	0.905	84.2	-0.008	0.423	0.482	80.0	-0.080	0.433	0.507	80.0
Bridge model A (BMa)	-0.128	0.325	0.424	88.9	-0.199	0.459	0.587	84.2	-0.099	0.244	0.302	95.0	-0.019	0.265	0.308	86.7
Bridge model B (BMb)	-0.070	0.322	0.440	87.0	-0.159	0.451	0.604	89.5	0.063	0.246	0.322	95.0	-0.074	0.251	0.301	80.0
Bridge model C (BMc)	-0.057	0.331	0.429	92.6	-0.286	0.455	0.570	89.5	0.174	0.250	0.300	100.0	-0.017	0.310	0.358	86.7
Bridge model D (BMd)	-0.083	0.333	0.439	88.9	-0.166	0.463	0.604	89.5	-0.018	0.265	0.340	95.0	-0.005	0.269	0.311	80.0
1st GDP outturn	0.035	0.216	0.332	94.4	0.057	0.389	0.513	89.5	0.055	0.152	0.201	100.0	0.040	0.207	0.244	86.7
Comparison with 1st outturn [2]	Ratio	DM	GW	PL	Ratio	DM	GW	PL	Ratio	DM	GW	PL	Ratio	DM	GW	PL
Random walk model (RW)	1.972	0.000	0.003	27.8	1.684	0.010	0.066	21.1	2.859	0.001	0.030	45.0	1.772	0.044	0.052	33.3
ARIMA model	2.049	0.000	0.003	22.2	1.733	0.011	0.073	31.6	2.888	0.031	0.058	20.0	1.982	0.048	0.069	40.0
Leading indicators A (LIa)	2.109	0.003	0.001	20.4	1.907	0.001	0.024	26.3	2.898	0.025	0.012	15.0	1.847	0.031	0.055	20.0
Leading indicators B (LIb)	1.807	0.000	0.005	25.9	1.522	0.017	0.123	47.4	2.494	0.017	0.034	10.0	2.277	0.000	0.012	26.7
Leading indicators C (LIc)	2.047	0.001	0.004	22.2	1.845	0.001	0.053	26.3	2.735	0.001	0.023	20.0	1.952	0.152	0.108	40.0
Leading indicators D (LId)	1.943	0.001	0.000	18.5	1.765	0.031	0.054	31.6	2.404	0.000	0.000	10.0	2.079	0.000	0.012	13.3
Bridge model A (BMa)	1.278	0.134	0.257	31.5	1.146	0.551	0.672	42.1	1.505	0.157	0.114	45.0	1.265	0.054	0.120	26.7
Bridge model B (BMb)	1.327	0.026	0.120	35.2	1.177	0.379	0.477	47.4	1.605	0.207	0.121	40.0	1.235	0.336	0.355	46.7
Bridge model C (BMc)	1.292	0.126	0.209	35.2	1.111	0.680	0.693	47.4	1.495	0.111	0.109	35.0	1.468	0.187	0.143	33.3
Bridge model D (BMd)	1.323	0.020	0.110	33.3	1.178	0.366	0.465	47.4	1.695	0.119	0.070	40.0	1.274	0.136	0.166	33.3

([1]) See the corresponding note in Table 5. ([2]) Forecasting ability comparison with respect to the benchmark (1st GDP outturn); see also the corresponding note in Tab. 5.

Table 7

FORECASTING ABILITY OF THE LATEST AVAILABLE GDP VINTAGE [1]

	Overall (91Q2-04Q3)				1st & 2nd blocks (91Q2-00Q4)				1st block (91Q2-95Q4)				2nd block (96Q1-00Q4)			
Forecasting models	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS	ME	MAE	RMSE	PS
Random walk model (RW)	-0.202	0.475	0.646	74.1	-0.171	0.528	0.719	76.9	-0.337	0.585	0.767	73.7	-0.013	0.474	0.670	80.0
ARIMA model	-0.133	0.487	0.656	75.9	-0.062	0.525	0.715	79.5	-0.243	0.584	0.764	73.7	0.109	0.469	0.665	85.0
Leading indicators A (LIa)	-0.179	0.497	0.661	79.6	-0.204	0.587	0.752	79.5	-0.343	0.708	0.873	73.7	-0.072	0.472	0.616	85.0
Leading indicators B (LIb)	-0.014	0.426	0.572	79.6	0.062	0.452	0.617	82.1	0.015	0.507	0.660	84.2	0.106	0.400	0.572	80.0
Leading indicators C (LIc)	0.041	0.484	0.643	83.3	0.087	0.560	0.716	84.6	-0.010	0.721	0.823	79.0	0.180	0.408	0.596	90.0
Leading indicators D (LId)	-0.018	0.467	0.585	79.6	0.031	0.522	0.645	82.1	0.014	0.656	0.759	79.0	0.047	0.395	0.513	85.0
Bridge model A (BMa)	-0.132	0.359	0.454	85.2	-0.150	0.399	0.503	87.2	-0.261	0.509	0.610	79.0	-0.044	0.293	0.372	95.0
Bridge model B (BMb)	-0.073	0.377	0.477	85.2	-0.047	0.424	0.529	92.3	-0.221	0.561	0.651	89.5	0.118	0.293	0.379	95.0
Bridge model C (BMc)	-0.061	0.386	0.483	90.7	-0.052	0.431	0.525	92.3	-0.348	0.522	0.620	84.2	0.229	0.344	0.414	100.0
Bridge model D (BMd)	-0.087	0.375	0.470	88.9	-0.093	0.424	0.528	92.3	-0.229	0.583	0.664	89.5	0.037	0.274	0.353	95.0
1st GDP outturn	0.031	0.286	0.440	96.3	0.054	0.368	0.514	97.4	-0.006	0.458	0.640	94.7	0.110	0.282	0.357	100.0
Intermediate GDP outturn [3]	n.a.	n.a.	n.a.	n.a.	-0.002	0.301	0.433	97.4	-0.062	0.380	0.558	94.7	0.055	0.225	0.263	100.0
Comparison with 1st outturn [2]	Ratio [3]	DM [3]	GW [3]	PL [3]	Ratio	DM	GW	PL	Ratio	DM	GW	PL	Ratio	DM	GW	PL
Random walk model (RW)	1.469	0.080	0.070	37.0	1.397	0.160	0.139	46.2	1.199	0.621	0.541	47.4	1.876	0.034	0.077	45.0
ARIMA model	1.491	0.036	0.067	31.5	1.390	0.125	0.160	41.0	1.195	0.542	0.532	42.1	1.861	0.097	0.144	40.0
Leading indicators A (LIa)	1.502	0.095	0.039	25.9	1.462	0.123	0.063	30.8	1.364	0.324	0.238	31.6	1.725	0.094	0.071	30.0
Leading indicators B (LIb)	1.301	0.137	0.166	35.2	1.199	0.356	0.380	43.6	1.032	0.913	0.904	52.6	1.602	0.118	0.169	35.0
Leading indicators C (LIc)	1.462	0.077	0.047	31.5	1.391	0.134	0.100	35.9	1.286	0.398	0.308	31.6	1.669	0.097	0.138	40.0
Leading indicators D (LId)	1.330	0.062	0.068	22.2	1.254	0.163	0.172	28.2	1.187	0.442	0.426	26.3	1.436	0.087	0.095	30.0
Bridge model A (BMa)	1.031	0.893	0.870	29.6	0.977	0.923	0.906	41.0	0.954	0.899	0.856	42.1	1.043	0.843	0.856	40.0
Bridge model B (BMb)	1.085	0.599	0.570	31.5	1.028	0.865	0.853	38.5	1.017	0.945	0.928	36.8	1.062	0.748	0.778	40.0
Bridge model C (BMc)	1.098	0.583	0.503	37.0	1.020	0.915	0.895	43.6	0.969	0.906	0.874	42.1	1.159	0.067	0.306	45.0
Bridge model D (BMd)	1.069	0.680	0.641	27.8	1.026	0.880	0.864	35.9	1.039	0.878	0.840	31.6	0.987	0.953	0.953	40.0
Intermediate GDP outturn [3]	n.a.	n.a.	n.a.	n.a.	0.841	0.171	0.231	59.0	0.873	0.450	0.448	57.9	0.737	0.013	0.069	60.0

([1]) See the corresponding note in Table 5. ([2]) Forecasting ability comparison with respect to the benchmark (1st GDP outturn); see also the corresponding note in Table 5. ([3]) Whole-sample statistics are not available because intermediate GDP outturn statistics were not computed for the 3rd block (where intermediate outturn data, by definition coincide with the latest available vintage).

Table 8

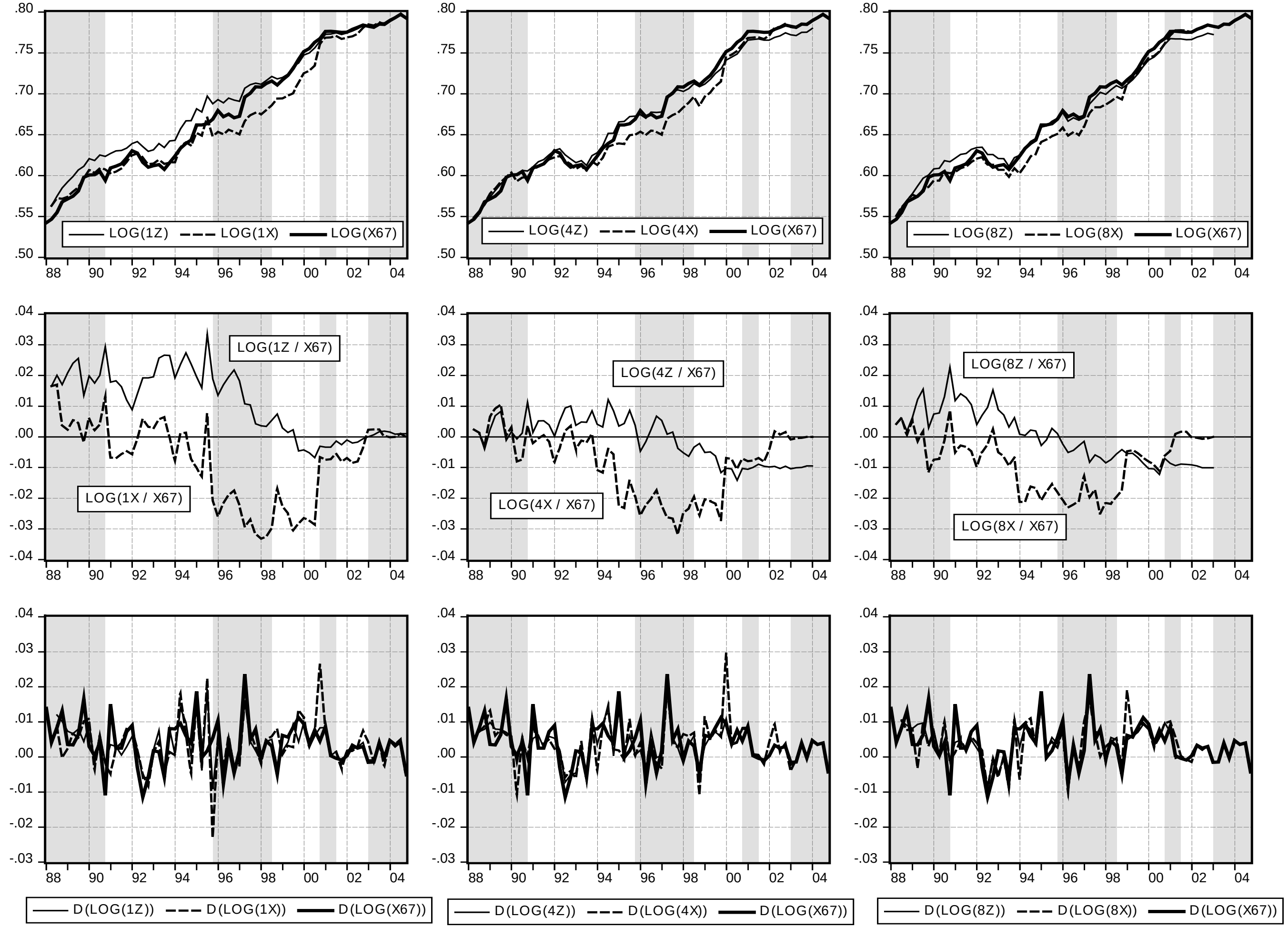
COMPARING INFORMATION OF 1ST OUTTURN AND MODEL-BASED FORECASTS [1]

Forecasting models [2]	with OLS standard .errors 1st outturn	with OLS standard .errors model-based	with White's standard errors 1st outturn	with White's standard errors model-based	with Newey-West s.e. 1st outturn	with Newey-West s.e. model-based
	Dependent variable: intermediate GDP outturn [3]					
ARIMA model	11.14	0.63	7.02	0.69	8.41	0.98
Leading indicators A (LIa)	10.82	-0.10	7.09	-0.09	8.26	-0.09
Leading indicators B (LIb)	10.11	-0.16	8.02	-0.12	10.05	-0.10
Leading indicators C (LIc)	10.83	0.02	7.58	0.02	9.47	0.02
Leading indicators D (LId)	10.65	-1.13	8.76	-0.79	11.85	-0.75
Bridge model A (BMa)	5.98	3.00	3.11	2.12	4.35	2.83
Bridge model B (BMb)	5.94	1.76	3.97	1.67	5.53	2.14
Bridge model C (BMc)	6.42	3.07	3.32	1.85	3.84	2.06
Bridge model D (BMd)	5.60	1.96	3.59	1.82	5.12	2.37
	Dependent variable: latest available GDP vintage [4]					
ARIMA model	7.18	0.86	3.69	1.17	3.47	1.18
Leading indicators A (LIa)	6.77	1.02	3.55	0.94	3.27	0.81
Leading indicators B (LIb)	6.10	0.93	3.26	0.78	3.23	1.10
Leading indicators C (LIc)	6.80	1.10	3.61	1.04	3.32	1.21
Leading indicators D (LId)	6.07	1.03	3.42	0.85	3.38	1.07
Bridge model A (BMa)	2.97	3.15	1.64	2.35	1.42	2.28
Bridge model B (BMb)	3.10	2.12	1.90	1.89	1.74	2.07
Bridge model C (BMc)	3.58	2.53	2.24	2.32	1.84	2.30
Bridge model D (BMd)	2.65	2.60	1.56	2.14	1.44	2.22

([1]) Fair and Shiller (1990) test over the 1991 Q2 – 2004 Q3 period. t-statistics (H_o: the corresponding parameter is zero) are obtained from the OLS estimation of β (1st outturn) and γ (model-based) parameters in equation (1) using different standard error estimators: OLS, White (1980), and Newey and West (1987). ([2]) Along the rows are listed the different models on which the forecasts are based (regressor "model-based"). ([3]) The "actual" GDP growth in equation (1) is measured by the intermediate outturn. ([4]) The "actual" GDP growth in equation (1) is measured by the latest available vintage.

Fig. 1

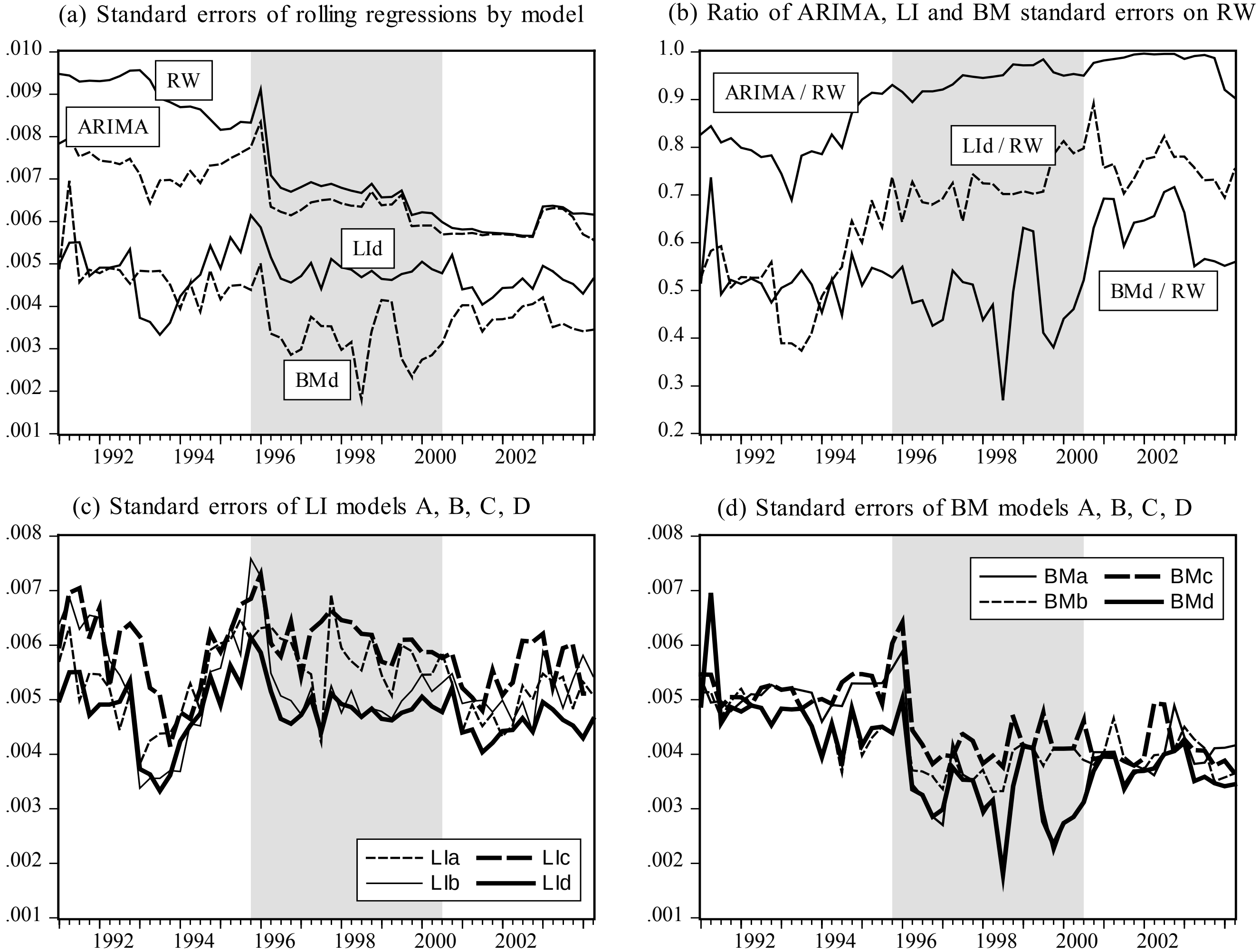
ALTERNATIVE REAL-TIME GDP SERIES FOR ITALY [1]



([1]) The first row shows GDP log-levels: 1X and 1Z are alternative 1st outturns (o=1), 4X and 4Z are the 4th outturns (o=4), 8X and 8Z are the 8th outturns (o=8), X67 is the latest-available (*i.e.* 67th) vintage. The middle row shows the total revisions from the 1st (4th, 8th) outturns to the latest available vintage (f=67). Finally, the plots in the last row show the corresponding (alternative) GDP quarterly growth rates by using the first-difference of log-levels: D(log(...)). Shading gives prominence to the periods between two consecutive benchmark revisions, see also Table 2.

Fig. 2

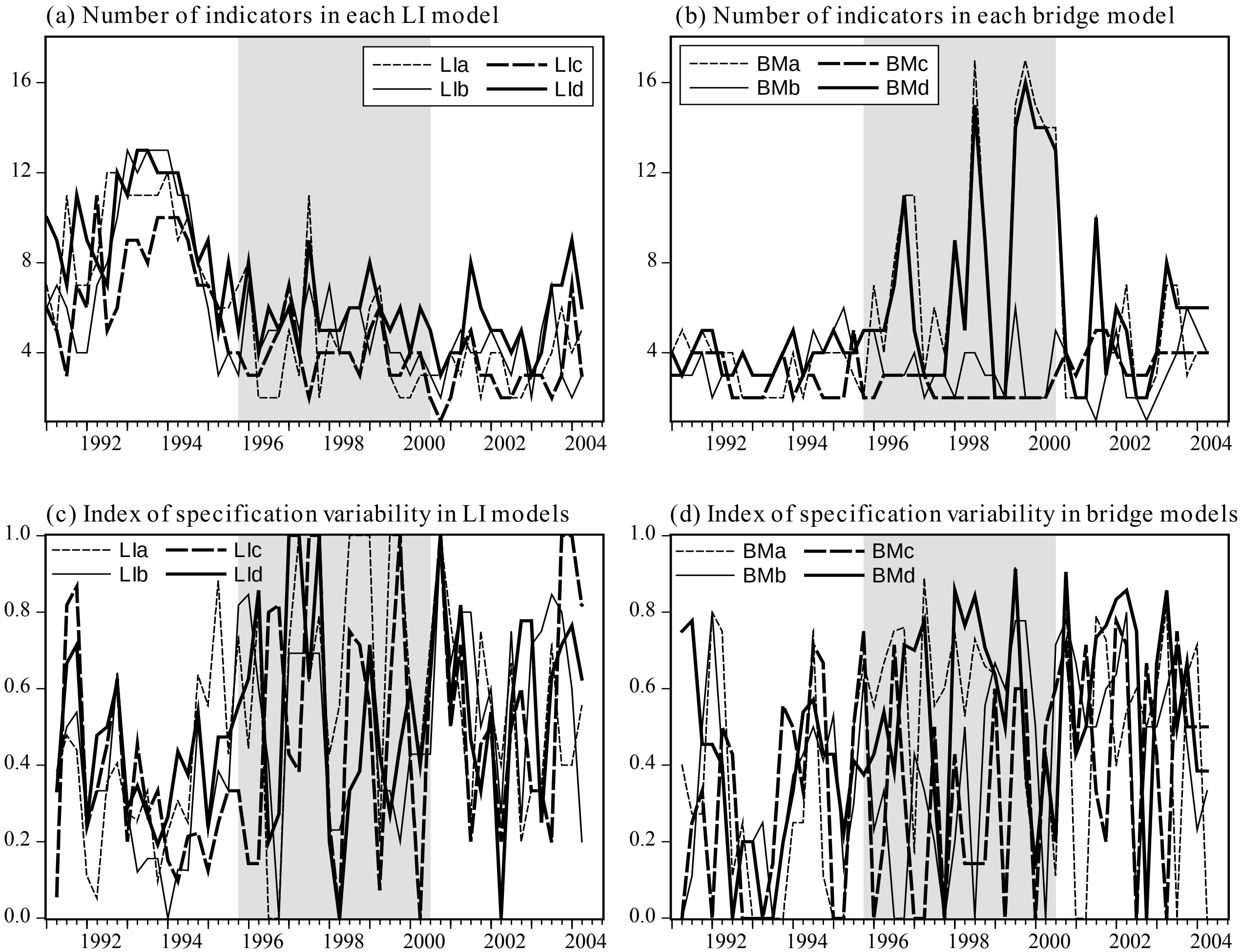
Comparison of the fitting performance of alternative models [1]



([1]) The standard errors of 530 rolling regressions are reported. Along the horizontal axes, the last observation of each regression is reported (the first observation being 80 quarters earlier). The shaded areas in the middle of each plot mark the end-periods of the rolling regressions belonging to the second block of GDP vintages.

Fig. 3

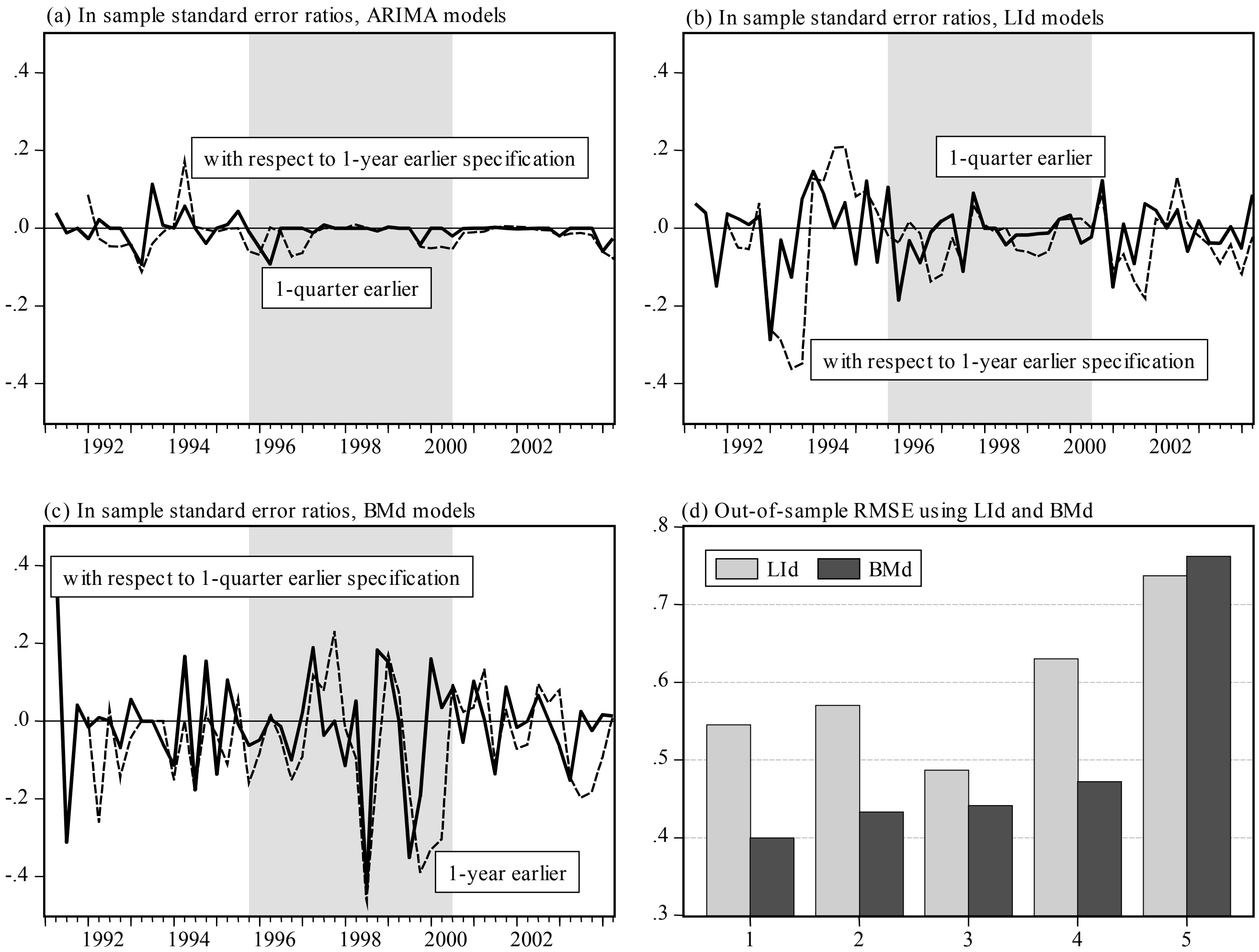
The specification variability of alternative models [1]



([1]) Along the horizontal axes, the last observation of each regression is reported (the first observation being 80 quarters earlier). The shaded areas in the middle of each plot mark the end-periods of the rolling regressions belonging to the second block of GDP vintages. The index of specification variability goes from 0 (when the model specification of vintage *v* is unchanged with respect to *v-1*) to 1 (when two subsequent specifications have no regressor in common).

Fig. 4

Analysis of the advantages of updating specification and estimation from one vintage to the other [1]



([1]) Plots (a), (b), (c): horizontal axes and shaded areas are described in Fig. 3; the ratios are s.e./s.e.$_1$-1 and (s.e./s.e$_4$)-1, where s.e. is the standard error of the regression (see Fig. 2), s.e.$_1$ and s.e.$_4$ are obtained from earlier specifications (of previous-quarter and previous–year vintages). Negative ratios are the cost of not updating the specification when a new vintage is available. Histogram (d): 1 (new specification and estimation for each vintage); 2-3 (only new estimation of respectively 1-quarter and 1-year earlier specifications); 4-5 (no re-estimation of 1-qrt and 1-y earlier specifications).