istat working papers

N.5
2023

A new hybrid framework to monitor business cycle: the RAT-Ita approach

Fabio Bacchini, Roberto Golinelli, Roberto Iannaccone, Davide Zurlo

Istat
Istituto Nazionale di Statistica

istat working papers

N.5
2023

A new hybrid framework to monitor business cycle: the RAT-Ita approach

Fabio Bacchini, Roberto Golinelli, Roberto Iannaccone, Davide Zurlo

Istat Working Papers

A new hybrid framework to monitor business cycle: the RAT-Ita approach

N. 5/2023

ISBN 978-88-458-2113-4

A new hybrid framework to monitor business cycle: the RAT-Ita approach

Fabio Bacchini[1], *Roberto Golinelli*[2], *Roberto Iannaccone*[1], *Davide Zurlo*[1]

Sommario

L'obiettivo del presente lavoro è quello di introdurre un nuovo indicatore del ciclo economico italiano, denominato RAT(ing)-Ita, in grado di monitorare mensilmente l'andamento del Pil. Per ogni mese, la metodologia proposta permette di stimare un valore nell'intervallo tra 0 e 1 che esprime la possibilità che il segno del Pil nel trimestre di riferimento sia positivo o negativo. La stima di RAT-Ita è articolata in tre fasi: selezione trimestrale degli indicatori che hanno una maggiore concordanza con gli andamenti del Pil (l'insieme degli indicatori congiunturali considerati è 1.285); nowcasting mensile, per ciascun indicatore, del segno del Pil per il trimestre di riferimento; aggregazione delle previsioni univariate. Il primo passo si basa sull'uso congiunto del test DAC (Directional Accuracy Change), della curva ROC (Receiver Operating Characteristic) e della coerenza spettrale. La stima univariata del segno del Pil è basata sull'utilizzo di modelli logit bivariati e la previsione dicotomica basata sulla curva ROC. Infine, nella fase di aggregazione si confronta la performance associata a diversi sistemi di ponderazione. La performance di RAT-Ita è stata valutata tramite un esercizio pseudo-real time per il periodo dal primo trimestre 2014 al terzo trimestre del 2022. Con riferimento al nowcasting del segno del Pil, RAT-Ita mostra risultati più accurati rispetto ai tradizionali modelli previsione.

Parole chiave: Nowcasting del segno del Pil; test DAC; curva ROC.

Abstract

The aim of this paper is to introduce a new business cycle indicator that exploits a large information set (1,285 time series) to nowcast the GDP evolution in real time. The proposed indicator, called RAT(ing)-Ita, focusses on the binary event represented by the sign of the one-step-ahead GDP growth rate. In other terms, we use many monthly indicators to "rate" the next-period economic performance represented by the GDP direction of change. The proposed methodology is organised along three steps: selection of the indicators, prediction of the GDP sign based on the single time series and aggregation of the single signal. The first step relies on the joint use of Directional Accuracy Change test, Receiver Operating Characteristic and spectral coherence. In the second step, the probability of the event sign of the GDP change, delivered by each selected indicator, is accomplished by using either bivariate logit models or the binary point prediction based on the ROC. Finally, in the aggregation step we adopt alternative weighting schemes. The performance of the methodology has been tested by predicting the Italian GDP quarter-on-quarter directional changes in pseudo-real time from Q1-2014 up to Q3-2022. The results are compared with the traditional benchmark models used to forecast the GDP,showing a better performance of RAT-Ita.

Keywords: GDP nowcasting with indicators, DAC test, ROC and spectral coherence.

JEL classification codes: C53, D31, E62.

1 Fabio Bacchini (bacchini@istat.it); Roberto Iannaccone (iannacco@istat); Davide Zurlo (zurlo@istat.it), Italian National Institute of Statistics/ Istituto Nazionale di Statistica – Istat.

2 Roberto Golinelli (roberto.golinelli@unibo.it), Alma Mater Studiorum – Università di Bologna, Italy.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this Istat working papers N. 5/2023.

Contents

Pag.

1. Introduction[3]

The interest in timely assessments of the business cycle evolution increases during uncertain times, like the ones we are facing since March 2020, characterised by *COVID-19* pandemic lock-downs and disruptions of the world supply chain, and by energy shocks fueled by the Ukrainian war.

The key point for a reliable short-run assessment is a large and timely information set, made of many indicators covering all aspects of economic activity.

However, such a large set of time series entails the issue of extracting valid signals from noisy indicators.

In order to track the short-run economic developments, several approaches are

available. Let's consider three of them:

- the nowcasting models that bridge the GDP evolution with related high frequency indicators (see, among the others, Baffigi *et al.* (2004), Kuzin *et al.* (2011), Bańbura *et al.* (2013)), and Babii *et al.* (2022));
- the construction of coincident and leading composite indexes of economic activity (see, among the others, Stock and Watson (1989), Aruoba *et al.* (2009), Altissimo *et al.* (2010), Cubadda *et al.* (2013), Proietti *et al*. (2021), and Lewis *et al.* (2022));
- models that relate the probability of recession to specific indicators, such as the yield curve spread or the subjective probability forecasts of the Survey of Professional Forecasters (see, among the others, Estrella and Hardouvelis (1991), Stock and Watson (1992), and Lahiri and Wang (2006)).

The bridge approach produces quantitative outcomes of the future economic activity represented by its most comprehensive measure (GDP). In this context, the interest focusses on the intensity of the next few changes in the target variable.

The composite index approach, rooted in the seminal project developed at the NBER (Mitchell and Burns, 1938), focusses on the identification of indicators able to summarise and forecast the state of the macroeconomic activity.

The probability of recession approach focusses on predicting recessions with indicators, rather than obtaining quantitative measures of the future economic activity.

Despite following different methodologies, the first and the second approaches share the same aim: tracking period by period the economic activity by predicting its future realisations. Instead, the third approach focusses on estimating the probability of occurrence of a particular outcome (in this case a recession).

The aim of this paper is to extend the toolbox of the short-run analyst (based on the three aforementioned approaches) with a new hybrid approach that exploits 1,285 indicators to

3 The authors would like to thank Gianluca Cubadda, Marco Lippi, Tommaso Proietti, and the other participants in the internal seminar at the Italian National Institute of Statistics - Istat (6th September 2022), and in the *2nd Workshop on Time Series Methods for Official Statistics* (22nd September 2022).

nowcast the GDP evolution. In particular, it nowcasts the binary event represented by the sign of the one-step-ahead GDP growth rate (*i.e.* the direction of change in GDP) and not its future realisations. In this, our proposal is closer to the third approach above (see also Lahiri and Wang, 2006), and is in line with recent works that aims to use together different methodology (see Kim *et al*., 2023). In other terms, we use many indicators to "rate" the next-period economic performance represented by the GDP direction of change.

Focussing on directional changes, our index has the advantage of providing robust predictions in unconventional times, when the economic system is subject to sudden and relevant shocks like those occurred since 2020 (for an alternative approach, see Barbaglia *et al*., 2022).

The methodological framework that we propose is organised along the following steps.

First, we select the subset of those indicators that better match the shortrun GDP directional changes up the the latest quarter available, *i.e.* that are the most conformed with the target. To do so, we jointly use the Directional Accuracy Change (Pesaran and Timmermann, 1992), the Receiver Operating Characteristic (ROC; Lahiri and Yang, 2018), and the spectral coherence with GDP growth over the cyclical frequency (2-8 years). The joint use of three alternative procedures relies on the idea that, although closely related, they can deliver different subset of selected indicators (Yang *et al*., 2023). As soon as a new GDP quarterly update is released, the list of the *best predictors* is reviewed. The ability of the selection to shrinking the large information set in fewer indicators is assessed against the evidence coming from the application of a dynamic factor model (see, among the others, Forni *et al*. (2001), Stock and Watson (2002a), and Forni *et al*. (2018)).

Second, we estimate the probability of the event *sign of the GDP change* delivered by each selected indicator either with bivariate logit models, or with the binary point prediction based on the ROC.

Third, we aggregate the single outcomes (related to each indicator) by using alternative weighting schemes presented in the literature (for a survey, see Lahiri and Wang, 2013).

The three-steps methodology has been applied to the Italian economy to realise a monthly indicator, that returns the expected sign of the GDP growth along the months of the current quarter. The monthly indicator lays in the $0-1$ range, where 0 means "positive change" and 1 "negative change".

Besides the estimate of the nowasting index, two by-products of our procedure are available:

a. a large and timely real-time databank of monthly indicators for Italy, in analogy with FRED-MD and FRED-QD for the USA (see McCracken and Ng (2016) and McCracken and Ng (2020)), with Coulombe *et al.* (2021) for the UK, and with Fortin-Gagnon *et al.* (2022) for Canada;
b. the time-varying list of the best predictors which is released each quarter.

The paper is organised as follows. Section 2 introduces the list and the main features of the 1,285 short-term indicators, and section 3 illustrates in detail the steps in which our methodological approach is organised. Section 4 shows the pseudo-real time exercise over a rolling window of 10 years (starting from the sample Q1-2004/Q1-2014), and focusses on the interpretation of the selection process using factor models, and through the comparison of the results with those of three traditional benchmark models (ARIMA, bridge and Midas).Finally, section 5 discusses future research to improve the approach of this paper.

2. Data

In recent years, the release of short-term indicators by national statistical agencies is getting better in terms of both coverage and timeliness. Researchers have particularly enjoyed this enrichment because the pandemic crisis and the Russian invasion of Ukraine have fuelled the need for real-time monitoring of the business cycle using techniques able to extract timely assessments about short-run evolution of economic activity (see for example Aprigliano *et al.*, 2023).

Recent literature has introduced different methodologies to exploit the information of large dataset of indicators. For example, Stock and Watson (2002b) refers to 149 monthly macroeconomic variables representing several different facets of the economy, while Forni *et al.* (2018) uses a dataset of 115 US macroeconomic and financial time series observed at monthly frequency, and Altissimo *et al.* (2000) analyses 183 time series (monthly and quarterly) to build up a leading indicator of the Italian economy published, up to March 2020, by the Italian national institute of statistics (Istat).

In the aforementioned studies, the target variable is often the first difference of the log of quarterly GDP, and the large datasets of indicators are organised in different blocks (such as prices, industrial production etc.).

In our work, we innovate on the previous literature in two respects. The first novelty regards the target variable (we model directional GDP changes), the second one regards the increase in the dataset size (we use more variables in each block and more blocks).

2.1 The target variable

In this paper, the target variable is the sign of the quarter on quarter (hereafter q-o-q) GDP growth. Therefore, we would need a monthly measure of GDP to exploit each update of the monthly indicators. However monthly GDP data are not available for almost all the countries. Temporal disaggregation of quarterly GDP data with monthly auxiliary indicators (such as industrial production) might be an option (see *e.g.* Chow and Lin (1971), Litterman (1983), and Fernández (1981)). However, although benchmarking techniques are now widely used in the national accounts practice, their use might generate monthly patterns of GDP too strictly related to those specific auxiliary indicators which are *ex ante* assumed to embody the unobservable monthly GDP fluctuations.

We propose to build up a monthly GDP series by repeating three times - for each month of the same quarter - the q-o-q GDP growth (see the example in Figure 2.1 related to the Italian GDP).

Figure 2.1 – The proposed scheme for the monthly GDP growth

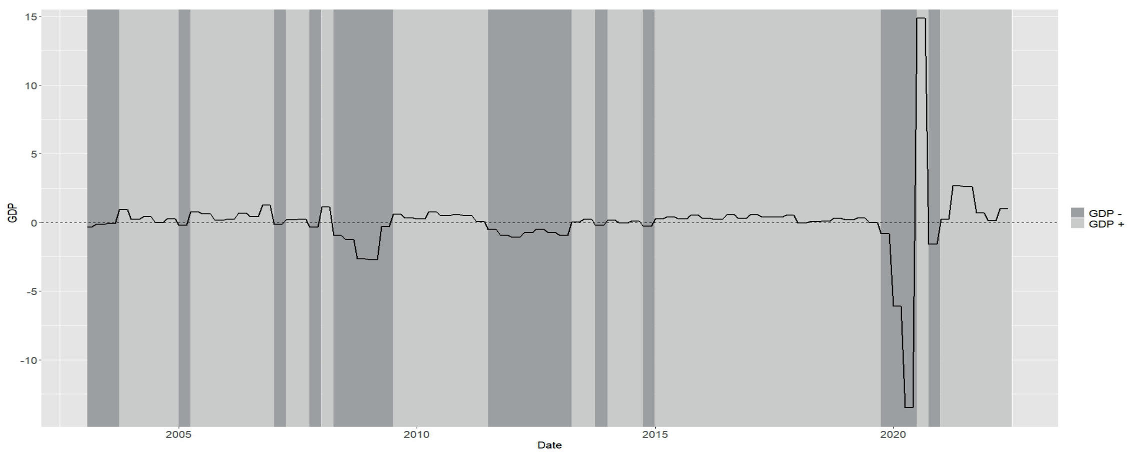
Q3-2021	2.59
Q4-2021	0.69
Q1-2022	0.13

jul-21	2.59
aug-21	2.59
sep-21	2.59
oct-21	0.69
nov-21	0.69
dec-21	0.69
jan-22	0.13
feb-22	0.13
mar-22	0.13

Source: Authors' elaboration on Istat data

Subsequently, the monthly GDP growth rates were transformed into a dichotomous variable (GDP_{01}) equal to 1 when the values are zero or negative, and equal to 0 when positive (respectively reported as grey and white areas in Figure 2.2[4]).

Figure 2.2 – The areas denoting the signs of q-o-q GDP growth



Source: Authors' elaboration on Istat data

This dichotomous monthly series of GDP directional changes is the target variable used for the selection of the "best" monthly indicators. Of course, the assumption leading to Figure 2.1 has an impact on the selection process, as it will tend to exclude those (transformed) indicators with higher volatility. For this, we used smoother data transformations (like third-differences, as shown below in this section), and exploited a mixture of selection approaches with different degrees of sensitivity to indicators' monthly volatility (see section 3.1).

4 This is not the same as shading the phases of the GDP path denoting recessions by using a dating algorithm, see Bry and Boschan (1971), and Harding and Pagan (2002) for respectively monthly and quarterly time series.

2.2 The dataset of monthly indicators

The indicators' dataset is composed of nine blocks (see Table 2.1). The industrial production (block 2) and business climate indicators in manufacturing and service (blocks 3 and 8) include variables that are highly disaggregated by 2-3 digits branches. The external trade variables (block 7) includes imports from and exports to the ten main Italian commercial partners (2-digits branches), and is the most represented one (538 time series). Further, the labour market (block 1) includes variables such as extra-time hours, payroll subsidies[5], inactive and employed workers; the consumer survey (block 5) includes households' sentiment and expectation variables; the prices set (block 6) includes producer and consumer prices by branch and category; the financial market (block 4) includes monetary aggregates and interest rates. Finally, the international market and Italian macro (block 9), includes a miscellanea of indicators such as the Economic and Policy Uncertainty index (EPU, see Baker *et al.*, 2016), the index of industrial production for Germany, USA and Euro area, the electricity consumption, and the index of car registrations, which are often used in monitoring short-term developments (see *e.g.* Bulligan *et al.,* 2010). Overall, the dataset includes 1,285 monthly variables[6].

Table 2.1 – The indicators' block coverage

Block	Name	No. of variables	SA
1	Labour market	207	X
2	Index of Industrial Production (3 digits), general and migs	102	X
3	Business climate indicators in manufacturing (2 digits)	155	X
4	Financial Market	58	
5	Consumer Survey	17	
6	Prices	38	
7	External Trade (Export and Import Main countries 2 digits)	538	X
8	Business climate indicators of the service sector	159	
9	International Market and Italian Macro	11	X
Total		1.285	

Source: Authors' elaboration

In order to assess their ability to explain the target in real-time, the monthly indicators in Table 2.1 must be classified by their timeliness (*i.e.* by the lag they are released with respect to the reference period), and transformed to be statistically congruent with the stationary target (*i.e.* the seasonally adjusted GDP growth).

As far as indicators' timeliness is concerned, the Italian case is representative of the general European statistical system. Table 2.2 summarises the calendar of indicators' publication lags at mid-May (*i.e.* when the GDP Q2 nowcast is made just after the preliminary GDP Q1, in the eve of the publication of Istat's note about short-run developments of the Italian economy). Some indicators (such as the surveys on confidence climates and the indices of

5 Payroll subsidies refers to the Cassa Integrazione Guadagni (CIG) wage supplementation mechanism.

6 The list of all the indicators is available upon request. Data are updated and made available on Istat's website each quarter, jointly with the new RAT-Ita release.

consumer prices) are classified as simultaneous (lag = 0 in Table 2.2) because their updates are released at the end of the reference month, while the updates of other indicators (such as the series of industrial production and of foreign trade extra EU) lag behind of respectively one and two months (lag = -1 and lag = -2 in Table 2.2).

Table 2.2 – Data available for the Q2 nowcast. Mid May

GDP	I quarter (preliminary)	Lag with respect to the nowcast date
Index of Industrial Production	March	-1
Business climate indicators in manufacturing	April	0
Business climate indicators of the service sector	April	0
Consumer Survey	April	0
HICP	April	0
PPI	March	-1
Financial Market	April	0
External trade	February	-1 / -2
Labour Market	February	-1 / -2
International Market and Italian Macrol	March	-1

Source: Authors' elaboration on Istat press release calendar

Concerning the transformation, we rely mostly on the available seasonally adjusted data ("SA" in the last column of Table 2.1). When this is not the case the TRAMO-SEATS approach (Gómez and Marvall, 1996) is used[7].

Furthermore, to rule out of trends and high persistence in the data, the survey indicator levels are untrasformed, while the other time series are expressed in differences of their levels or log-levels[8]. Table 2.3 lists the data transformations by block. After treatment, the resulting time series are standardised.

Table 2.3 – Treatment of the time series

Domain Name	Treatment
Labour market	Extra time (in differences), Payroll subside (in log-dif ferences)
Index of Industrial Production	Log-differences
Business climate indicators in manufacturing (2 digits)	Levels
Financial Market	Loan (log-differences) Interest rates (in differences)
Consumer Survey	Levels
Prices	Log-differences
External Trade	Levels
Business climate indicators of the service sector	Levels
International Market and Italian Macro	Log-differences

Source: Authors' elaboration

7 For the seasonal adjustment has been use the package RJDemetra in R with the specification of automatic choice for the transformation of the series, the estimation of the ARIMA model and the outliers identification.

8 In particular, we refer to the difference between month t and month *t*-3. The choice of treating the variables in this way (instead of in first-differences) aims to prevent the transformed series to be excessively volatile.

3. RAT-Ita: the Hybrid Approach

This paper aims to introduce a novelty approach to nowcast, in each month, the sign of the current q-o-q growth rate of GDP, not yet released. In other terms, we propose a sort of rating for the economy that, applied to Italy, is labelled as RAT(ing)-Ita.

The methodology is carried out by a three-step procedure:

1. selection of a time-varying subset of monthly series which picks the most suitable indicators to track the signs of GDP directional changes up to the latest available quarter;
2. estimation of the signal coming from each selected-indicator to nowcast the GDP sign;
3. aggregation of the signal on the GDP sign provided by each selected indicator.

The procedure runs each quarter, after the new GDP figure is released, providing the update of the selected indicators' list.

The selection step relies on the combination of two techniques related to the dichotomous nature of the target variable (GDP_{01}) - the Directional Accuracy test (DAC) and the Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC) - together with a third one, more traditional, based on the spectral coherence between the quarterly growth rate of GDP and of each indicator[9].

In the empirical application below, we show how the combination of different techniques improves the nowcast. This multi-technique procedure explains why we termed our approach as "hybrid".

3.1 Selection procedure

DAC is a non-parametric test introduced by Pesaran and Timmermann (1992) to evaluate the ability of an indicator to predict the sign of the target (in our case, the q-o-q GDP growth rate). Denoting the series of interest as y_t and with x_t its forecast, $t = 1, \dots, T$, the test statistic is defined as:

$$S_n = \frac{P - P_*}{[V(P) - V(P_*)]^{0.5}}$$

9 Besides these three, a fourth selection technique, LASSO, was used without improving the performance.

where:

$$\begin{aligned}
P =& T^{-1}\sum_{t=1}^{T} I(y_t x_t) \\
P_* =& P_y P_x + (1 - P_y)(1 - P_x) \\
V(P) =& T^{-1} P_*(1 - P_*) \\
V(P_*) =& T^{-1}(2P_y - 1)^2 P_x(1 - P_x) + T^{-1}(2P_x - 1)^2 P_y(1 - P_y) + \\
& 4T^{-2} P_y P_x(1 - P_y)(1 - P_x) \\
P_y =& T^{-1}\sum_{t=1}^{T} I(y_t) \text{ and } P_x = T^{-1}\sum_{t=1}^{T} I(x_t)
\end{aligned}$$

with:

$$I(\cdot) = \begin{cases} 1 & \text{if } \cdot > 0 \\ 0 & \text{otherwise} \end{cases}$$

Under the null hypothesis of x_t not being able to predict y_t, Sn follows the standard normal distribution.

AUC-ROC is a technique introduced in the field of medical statistics (Lusted, 1960). Recently, the AUC-ROC approach has been used in the economic fields of business cycle forecasting (Berge and Jordà (2011), and Lahiri and Wang (2013)), to determine firms' export threshold, (Costa *et al.,* 2019), and to binary forecasting evaluation (Lahiri and Yang, 2013).

The aim of AUC-ROC is to define the ability of a time series (classifier) to predict the 0 or 1 states of a target variable. According to Fawcett (2006), given the target y and the classifier x, four possible outcomes are feasible for each observation: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), as shown in Figure 3.1.

The ability for a classifier to track the target variable can be measured using two main metrics: sensitivity and specificity. Sensitivity represents the probability of detecting true positives $TP/(TP + FN)$, while specificity is the probability of detecting true negatives, *i.e.* $TN/(TN + FP)$. Specificity is usually analysed looking at its reciprocal expression $FP/(TN + FP) = (1 - \textit{specificity})$, which measures the probability of false positives. For each observation of the classifier x_t (with $t = 1, \ldots, T$), we can define:

$$\begin{aligned}
\text{sensitivity}(x_t) =& \; TP(x_t)/[TP(x_t) + FN(x_t)] = \; P(X \geq x_t | Y = 1) \\
\text{1-specificity}(x_t) =& \; FP(x_t)/[TN(x_t) + FP(x_t)] = \; P(X \geq x_t | Y = 0)
\end{aligned}$$

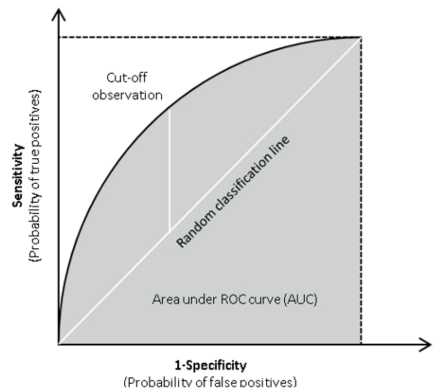
Figure 3.1 – The classification of states

		Estimated classification	
		1	0
True classification	1	*TP*	*FN*
	0	*FP*	*TN*

Source: Fawcett (2006)

The ROC curve in Figure 3.2 reports the possible combinations of *sensitivity* and (1 − *specificity*). The grey area below the ROC curve (AUC-ROC) measures the classifier ability to track the target indicator compared to the random classification along the 45° line. The AUC-ROC assumes values in the interval [0,1] and, when the area is less than 0.5, a totally random classification is better than the one associated to the classifier. Instead, when the area is equal to 1, the classifier x_t perfectly defines all the different $0 - 1$ states of the target.

Figure 3.2 – The ROC curve



Source: Authors' elaboration

We also used AUC-ROC to choose the best lag of the indicator (subject to its timeliness in Table 2.2), *i.e.* the lag corresponding to the best forecasting ability of that indicator.

DAC and AUC-ROC results are largely affected by the assumption that the sign of GDP directional changes in each month of a quarter is constant and equal to the sign of the q-o-q growth of that quarter (see Figure 2.1). Therefore, in the selection step, these techniques penalise the indicators with higher monthly volatility.

This issue can be mitigated by introducing a third technique of selection based on the spectral coherence[10] at business cycle frequency (2-8 years) between the quarterly GDP growth rates and the quarterly average of the monthly candidate-indicators in our dataset[11].

The best (selected) indicators' subset includes two types of variables: those with a strong monthly relationship with GDP_{01} and those with a strong cyclical relationship with the GDP quarterly growth rates[12]. This mixture of variables leads to better predictive results, as we will show in the section reporting the empirical results.

The full list of n selected monthly indicators is obtained by merging the single lists coming from each of the three techniques discussed in this section. The selection process is run each quarter, after a GDP update is released, while the selection is left unchanged for the following months until a new GDP release occurs.

3.2 Nowcasting different GDP signs and aggregation

Each month, the sign of the current-quarter GDP growth is estimated n times (one for each selected indicator) by using two alternative approaches:

- **the probability prediction** which estimates the conditional probability of occurrence of a negative sign for the target;
- **the binary point prediction** which transforms the selected indicator in a binary variable, delivering the positive/negative classification of the target (see Lahiri and Yang, 2013).

Each approach used to nowcast the GDP sign is characterised by 4 different aggregation methods. In this way we provide and compare 8 different values for the nowcasting of the GDP sign.

3.2.1 The probability prediction

Using logit models, each of the n selected time series is regressed against the GDP, expressed in term of 0 (positive quarterly growth rate) and 1 (negative quarterly growth rate):

$$GDP_{01} = \Lambda(c + x_i\beta_i) \; i = 1, \dots, n \tag{1}$$

where Λ is the logistic function distribution, and x_i is the $i-th$ selected time series with T observations. The model is estimated using a training dataset that includes the last available GDP quarter and the indicator according to both its timeliness (see Table 2.2), and its lag

10 The basic call is to the function mvspec, which is available in astsa package. To compute averaged periodograms we use the Daniell window. The notion of analysing frequency fluctuations using classical statistical ideas extends to the case in which there are several jointly stationary series.In this case, we can introduce the idea of a correlation indexed by frequency,called the coherence.

11 Of course, we could compute the indicator at quarterly frequency by averaging 3 months only when all the months were available. Regarding the average of the last quarter for those indicators released at lag=2, we considered as the quarterly average only the first (only available) month, while the quarterly average for the indicators available at lag = 1 was computed by considering only the first two available months of the quarter.

12 Below, in the applied parts, we will introduce and discuss the definition of specific thresholds to qualify an indicator as "strongly" related with the target.

with the highest AUC-ROC. Each logit regression returns the estimated probabilities $\hat{p}_i$ for the event GDP_{01}.

The *n* probability values can be then aggregated using different aggregation methods. Starting from the simplest, we have the arithmetic mean (henceforth **Amean**):

$$\hat{p}_A = \frac{\sum_{i=1}^{n} \hat{p}_i}{n} \tag{2}$$

A second aggregation method has been implemented to compute the corrected geometric mean (henceforth **Gmean**):

$$\hat{p}_G(a) = \frac{\left[\prod_{i=1}^{n} (\frac{\hat{p}_i}{1-\hat{p}_i})^{1/n}\right]^a}{1 + \left[\prod_{i=1}^{n} (\frac{\hat{p}_i}{1-\hat{p}_i})^{1/n}\right]^a} \tag{3}$$

where $a \geq 1$ is an unknown parameter used in Satopää *et al.* (2014) to correct the systematic bias in the forecast probability. In particular, the parameter a is estimated through the following minimisation problem:

$$\hat{a} = \arg\min_a \sum_{t=1}^{T} (\hat{p}_{G,t}(a) - GDP_{01t})^2 \tag{4}$$

where $\hat{p}_{G,t}$ is the aggregate probability forecast for *t*, and the event indicator $GDP_{01t} \in \{0, 1\}$ depends on whether the growth rate at time *t* is negative ($GDP_{01t} = 1$) or positive ($GDP_{01t} = 0$).

A third aggregation method is the *logarithm opinion pooling*:

$$\hat{p}_L = \frac{\prod_{i=1}^{n} \hat{p}_i^{w_i}}{\prod_{i=1}^{n} \hat{p}_i^{w_i} + \prod_{i=1}^{n} (1 - \hat{p}^i)^{w^i}} \tag{5}$$

where we assume equal weights for each probability ($w_i = 1/n$) to obtain an equally-weighted logarithmic opinion pool (henceforth **Elop**).

The fourth aggregation method is the *Beta-transformed linear opinion pool* (henceforth **Beta**):

$$\hat{p}_B(\alpha, \beta) = H_{\alpha,\beta}\Big(\sum_{i=1}^{N} p^i w^i\Big) \tag{6}$$

where $H_{\alpha,\beta}$ is the cumulative distribution function of the Beta distribution with parameters α and β, and with weight w_i (again, we assume equal weights to each probability). In line with the previous studies (Satopää *et al.* (2014), overfitting is controlled using $\alpha = \beta \geq 1$. The parameter $\alpha = \beta$ is estimated through the following log-likelihood maximisation (see Ranjan and Gneiting, 2010):

$$\hat{\alpha} = \arg\max_{\alpha} \sum_{t=1}^{T} GDP_{01t} \log\left[H_{\alpha,\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{p}_{it}\right)\right] + \sum_{t=1}^{T}(1 - GDP_{01t}) \log\left[1 - H_{\alpha,\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{p}_{it}\right)\right] \tag{7}$$

3.2.2 The binary point prediction

The other approach uses diffusion indexes. Each selected time series is first transformed in a binary variable, then the proportion of classifications equal to 1 is measured. When the diffusion indexes are above the threshold 0.5, the GDP growth rate can be interpreted to be negative, positive when below 0.5. In the benchmark method (henceforth **Bench**) each time series is classified as 0 or 1 when the growth rate is respectively positive or negative.

Another method is based on the use of the ROC approach. In this context, the idea is to identify a threshold c_i (one of the observed values of the *i−th* selected time series) which maximises the indicator's ability to meet the evolution of our target GDP_{01} (see again Figure 3.2).

The resulting threshold (c_i) is then compared with the last value of the selected indicator: if ci is lower we classify the GDP_{01} as 1 (as 0 if higher).

Different criteria have been used to define the threshold:

- a criterion based on Youden's Index (henceforth Youden) defined as (see Youden, 1950):

$$\max(\text{Sensitivity}(c_i) + \text{Specificity}(c_i)) \tag{8}$$

where the threshold is the observation c_i that maximises the equation;

- a criterion based on the minimum *p*-value relative to the Chi-squared test (henceforth **Chi.sq**), which measures the association between the sign of q-o-q GDP growth rate and the binary result obtained using different thresholds (Miller and Siegmund, 1982).
- a criterion called *closest.topleft* (henceforth **CTL**) for which the optimal threshold is the observation that satisfies the condition:

$$\min((1 - \text{Sensitivity}(c_i))^2 + r \times (1 - \text{Specificity}(c_i))^2) \tag{9}$$

where r = (1 − prevalence)/(cost × prevalence), cost is the relative cost of a false negative classification (as compared with a false positive classification), and prevalence, or the proportion of cases in the population, is the mean of GDP growth rate sign up to the previous quarter).

4. The pseudo-real time nowcast

The three-steps approach described above has been applied to the Italian case to estimate the monthly nowcasting indicator for the sign of the q-o-q GDP growth rate. In doing so, we explore the performance of the approach with reference to the calibration of the thresholds of the criteria of selection and other parameters of interest.

The selection criteria run over a rolling window of 10 years starting from the sample Q1-2004/Q1-2014. For each quarter, the three-step procedure use the entire dataset available (1,285 time series) providing the nowcast of the GDP sign for each month in the period Q2-2014 to Q3-2022[13].

The pseudo-real time estimation includes both an assessment of the performance of the selection procedure by using, for a specific quarter, the dynamic factor model approach and a comparison of the results with the traditional benchmark models (AR, Brigde and Midas models).

4.1 Selection procedure: empirical results

The selection procedure requires specific thresholds for each criterium (DAC, AUC-ROC and spectral coherence).

As far as DAC is concerned, we set the 5 percent significance level; therefore *p*-values below 0.05 imply rejection of the null hypothesis that indicator information is not helpful to predict the GDP sign.

Regarding the other two criteria, AUC-ROC (with monthly data) and spectral coherence (with quarterly data), the single-target thresholds are: AUC-ROC > 0.79, and spectral coherence > 0.85. In the application we relies on the use of the following combination of the two thresholds:

- AUC-ROC > 0.79 & Spectral Coherence > 0.55 (AUC-ROC criterium);
- Spectral Coherence >0.85 & AUC-ROC > 0.65 (Coherence criterium).

To illustrate how the thresholds work, Table 4.1 reports the number of selected indicators in Q3-2022 using alternative combinations of thresholds for AUCROC and spectral coherence criteria. The total number of time series are 409 indicators (in 1,285) that reject at 5 percent the null hypothesis of the DAC test.

Only 6 indicators jointly match the AUC-ROC (> 0.79) and the spectral coherence (> 0.85) criteria. Instead, the introduction of lower thresholds for either AUC-ROC (> 0.65) or spectral coherence (> 0.55) criteria respectively select further 39 and 24 indicators, leading to a total selection of 69 series[14].

Table 4.1 – Alternative AUC-ROC and spectral coherence thresholds results for Q3-2022

AUC-ROC	Coherence		
	0 - 0.55	0.55 - 0.85	0.85 - 1
0 - 0.65	64	69	22
0.65 - 0.79	49	136	**39**
0.79 - 1	0	**24**	**6**

Source: Authors' elaboration

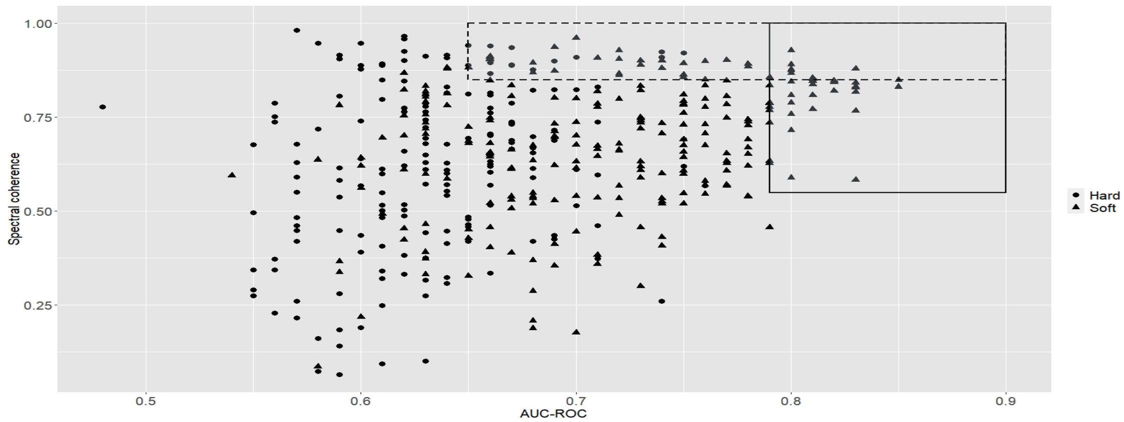
13 Recursive sample results, starting in Q1-2003, were also obtained, with results in line with those of the rolling window.

14 The proposed thresholds are set on the basis of the outcomes of alternative thresholds. A more comprehensive theoretical approach to derive the optimal values for thee threshold is in the agenda.

The AUC-ROC and spectral coherence criteria work heterogeneously between the blocks of the information set. Partly due to the assumption of a GDP sign constant for each month of the quarter, the AUC-ROC is much more oriented towards selecting business and consumer survey indicators, while the spectral coherence is much more oriented towards selecting hard indicators (*e.g.* industrial production, labour market, external trade, etc.).

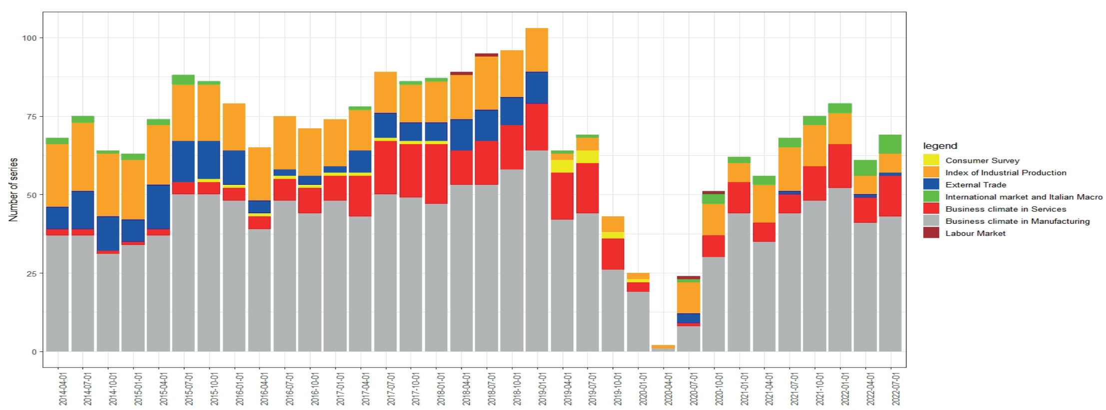
Figure 4.1 shows the effects of the two criteria on the composition of the selected indicators in Q3-2022 as an example. AUC-ROC (solid rectangle) selects only soft indicators (triangles), while spectral coherence (dashed rectangle) also selects hard indicators (dots).

Figure 4.1 – Performance of AUC-ROC and spectral coherence on selection for Q3-2022



Source: Authors' elaboration

Figure 4.2 – Selected indicators by block in each quarter



Source: Authors' elaboration

Given that the selection criteria run every quarter, Figure 4.2 plots for the main blocks the number of time varying subset of selected indicators according to the hybrid approach. Several empirical findings are related to the performance of the selection method. *COVID-19* introduces a drastic reduction in the number of selected time series. The relevance of the external trade block is high up to Q1-2019, then dramatically drops in the following quarters.

As illustrated in the example based on the dynamic factor model (see next paragraph), this result might be related to the high degree of details of the time series selected for the external trade that implies a large heterogeneity.

Overall, the blocks with the highest number of indicators are those of the manufacturing and service business surveys, because of their timeliness, together with the index of industrial production.

4.1.1 The selection procedure and the factors analysis

The proposed methodology is based on the exploitation of the relationship between the target variable (GDP_{01}) and each single monthly time series. However in the literature, the use of a large number of time series is often associated with the class of linear high dimensional Dynamic Factor Models (DFM).

In this paper, that follows a different methodological approach, we apply the DFM to assess on the quality of the selection process performed by RAT-Ita.

As pointed out by Forni *et al.* (2018), large-dimensional DFM represent each variable in the dataset as decomposed into a common component, driven by a small (as compared to the number of series in the dataset) and fixed (as the number of series increases) number of common factors, and an idiosyncratic component.

$$x_{it} = \chi_{it} + \xi_{it}$$

Avoiding to present the general representation of the dynamic factor model (see for example Forni *et al.* (2000) and Stock and Watson (2002a)), we assume that for *a* given *t* the common components x_{it}, for $i \in N$, span a finitedimensional vector space S_t. Stationarity of the common and idiosyncratic components implies that the dimension of St, call it r, is independent of t and there exists a *stationary basis* $Ft = (F_{1t}F_{2t} \ldots Frt)$ such that the so-called static representation of Stock and Watson (2002a) is defined as:

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \ldots + \lambda_{ir}F_{rt} + \xi_{it}$$

where the *r* factors F_{jt} and the loadings λ_{jt} represent the common component x_{it}. The idiosyncratic component ξ_{it} is assumed to be orthogonal across different variables or only weakly correlated, so that the covariance of the variables is mostly accounted for the common component. In the application, the factors and the loadings are estimated using the first *r* principal components.

Once the static factor model is estimated for each of the 9 blocks, for all the indicators considered as a whole (Total) and for the selected sub-sample (Selection), to assess the relative importance of the common component, we consider:

- the ratio of the variance of the common component on the total variance of the time series *i* (Boivin and Ng, 2006):

$$R_i^2 = \frac{\sum_{t=1}^{T} \hat{\chi}_{it}^2}{\sum_{t=1}^{T} x_{it}^2}$$

- the cross-section dispersion measured by the distributive measures of the ratio R^2_i in the 90 (p90) and the 10 percentile (p10).
- the correlation between the estimated factors across the blocks and those across the small number of selected indicators to account for their commonality.

To investigate the importance of the selection process in terms of the estimated common component we consider only a quarter of the pseudo real-time exercise. The selected time span runs from Q1-2003 to Q3-2019 where the number of the selected indicators is equal to 69. Additionally, the estimation of the static factor model is performed considering 3 factors ($r = 3$[15]).

The results in Table 4.2 suggest that the selection process of RAT-Ita leads to a variance ratios which are higher than those of the one measured on the full dataset of available indicators (first row, latest 2 columns). Even considering the comparison with each of the single block, the variance ratio of *Selection* is higher but with the exception of consumption surveys and financial indicators blocks, where, a very strong degree of correlation across the indicators in the block exists before the estimation of the common component. Even the dispersion (p90-p10) for the *Selection* (last row, latest column) appears as one of the lowest compared to the total number of the series. This measure is related to the details presented in the previous rows related to the value of the 10th percentile (second row), the median (third row) and the 90th percentile (four row).

Both the ratio of the variance of the common component and the dispersion measures return poor results for the block of external trade and the labour market. These evidences are in line with the ones provided by the selection process presented before (see Figure 4.2) that consider only a small set of time series belonging to the two blocks.

The DFM estimation enriches this evidence underlining how the time series belonging to the blocks are heterogeneous among them in terms of the common components. Expecially for the external trade block this could be related to the evidences presented by Boivin and Ng (2006) where different Data Generation Process for each time series implies a poorest estimation for the common component.

Table 4.2 – Ratio of the variance explained by the common component and dispersion measures for each block, Total and Selection

	External trade	Consumer survey	Financial market	Labour market	Int. market IT macro	Business manifacturing	Prices	IPI	Business service	Total	Selection
Var. ratio	0.11	0.74	0.83	0.17	0.49	0.57	0.37	0.25	0.34	0.25	0.68
P10	0.01	0.52	0.65	0.01	0.06	0.12	0.08	0.05	0.08	0.01	0.42
Median	0.07	0.74	0.88	0.09	0.49	0.66	0.26	0.21	0.32	0.15	0.71
P90	0.28	0.91	0.97	0.47	0.86	0.81	0.80	0.47	0.64	0.69	0.90
P90 - P10	0.27	0.39	0.33	0.46	0.80	0.68	0.72	0.42	0.56	0.68	0.48
N	**538**	**17**	**58**	**187**	**31**	**155**	**38**	**102**	**159**	**1285**	**69**

Source: Authors' elaboration

Further evidences based on the result of the DFM estimation support the quality of the selection process of RAT-Ita. Considering the estimation of the three factors for each block,

15 The number of the optimal factors could be derived using the test proposed by Bai and Ng (2002). In this application we set the number of the factors to 3 for all the blocks. Results does not change considering 4 factors that is the value identified by Aprigliano *et al.*, 2023 for the Italian economy.

for the Total and the Selection group a correlation between them has been elaborated[16]. The factors estimated on the Selection group shows an high degree of correlation with the factors estimated for the total set of time series and highly correlated with the business survey (manufacturing and services), prices and industrial production. Overall considering Q3-2019, the estimation of the DFM seems to confirm the high quality of the selection process based on RAT-Ita approach. The selected set of time series shows high performance in terms of the estimated factors and common components, capturing *the common engine* of the Italian business cycle.

4.2 Nowcasting GDP signs for each selected indicator

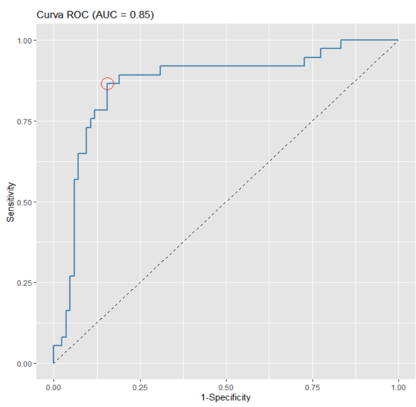
According to the proposed methodology, once an indicator has been selected a nowcast for the sign of GDP quarterly growth rate is needed. In order to reach this goal we use the binary point prediction methods shown above to identify a threshold or the conditional probability of negative GDP growth rate is estimated through a logit model (probability prediction).

We illustrate the approaches with an example based on the index of industrial production of the "Manufacturing of machinery and equipment not elsewhere classified" sector, selected within the industrial production block.

In the Figure 4.3 the blue line is the ROC curve, for which the AUC-ROC is equal to 0.85. The circle in red is the value of the threshold identified through the criterion *closest.topleft i.e.* the observation of series that satisfies equation 9.

The threshold value selected with the *closest.topleft* is then $c_i = 0.26$ (horizontal line in top panel of Figure 4.4). This threshold is used for the binary prediction point of the series to nowcast the GDP growth rate for the third quarter with data up to July. The time series predicts a positive GDP growth rate because the last value is higher than the threshold c_i.

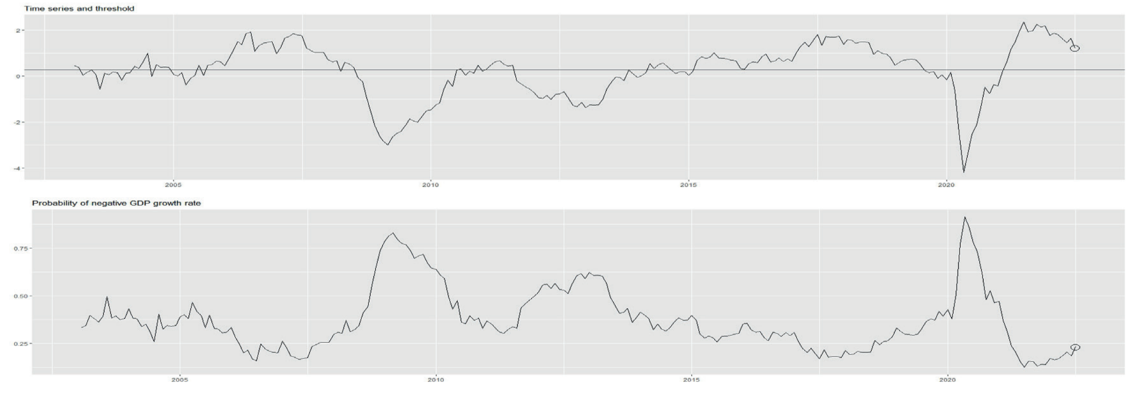
Figure 4.3 – The ROC curve in practice



Source: Authors' elaboration

16 The results are available on request.

Figure 4.4 – The classification in practice



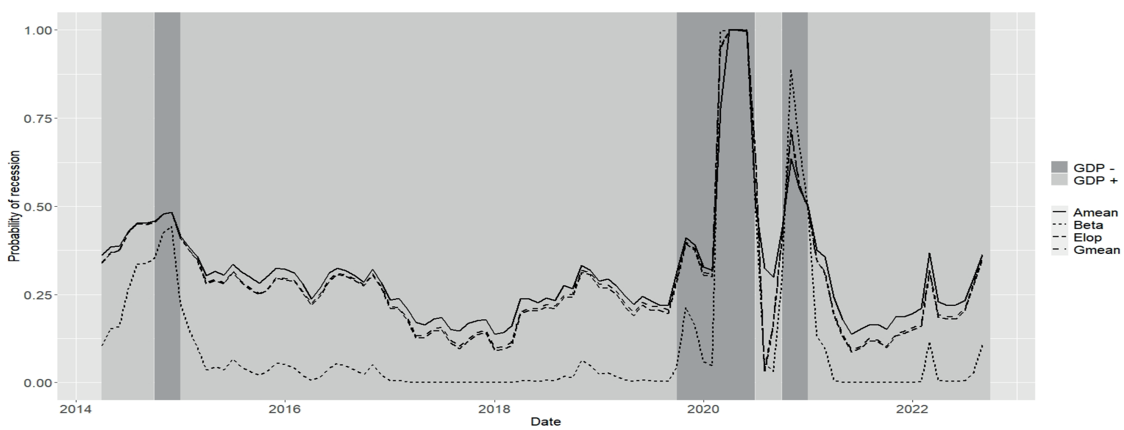
Source: Authors' elaboration

In the bottom panel of Figure 4.4, the fitted values of the logit model in equation 1 are shown together with the predicted probability (black circle) for September which indicates positive growth rate for the third quarter because the $P\ (GDP_{01Q3-22} = 1 \mid x^i) = 0.12$.

4.3 Aggregation and final nowcast of GDP sign

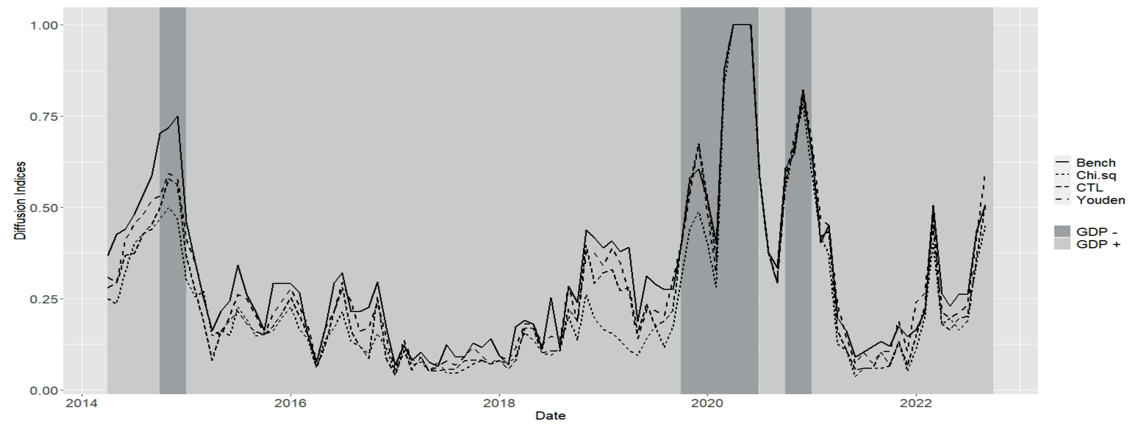
The procedure described in the previous paragraphs has been applied for each month in the time-span from April-2014 to September-2022. After the indicators' selections, 8 different monthly composite indicators can be obtained according to either the 4 aggregations of probability prediction in Figure 4.5, or the 4 aggregations of binary point prediction (diffusion indices) in Figure 4.6.

Figure 4.5 – RAT-ITA using probability prediction approach



Source: Authors' elaboration

Figure 4.6 – RAT-ITA using diffusion index approach



Source: Authors' elaboration

The performance of the different aggregation methods has been evaluated using the set of statistics proposed in Lahiri and Wang (2013), together with the correlation and the AUC-ROC between the GDP_{01} and the composite indicator.

Among the proposed statistics, we consider the Brier's Quadratic Probability Score (QPS), a probability analog of the mean squared error:

$$QPS(f, GDP_{01}) = \frac{1}{T}\sum_{t=1}^{T}(f_t - GDP_{01t})^2 \tag{10}$$

where f_t is the combined forecast probability or diffusion index of the event at time t; and GDP_{01t} is the realisation of the event (1 if the GDP growth rate is negative and 0 otherwise) and T is the number of observations. The QPS can assume only values between 0 and 1. A score of 0 corresponds to a perfect accuracy.

The estimated QPS values is usually compared with the one obtained defining a benchmark forecast. The most common benchmark forecast is the unconditional probability of the event of interest, known as the base rate calculated as the mean of GDP_{01} for the full period. We calculate the skill score (SS) measure as:

$$SS(f, GDP_{01}) = 1 - \frac{QPS(f, GDP_{01})}{QPS(\overline{GDP}_{01}, GDP_{01})} \tag{11}$$

where $PS(\overline{GDP}_{01}, GDP_{01})$ is the mean squared error associated with benchmark forecast $\overline{GDP}_{01}$ equal to 0.31 in our sample. A value of the skill score more close to 1 indicate substantial improvements over the benchmark base rate forecasts.

In addition to the measurements already described we also consider a test to evaluate the null hypothesis of no difference in the accuracy of two competing forecasts. In particular we implement the extension introduced in Lopez (2001) to the S_1 test of Diebold and Mariano (2002). According to Lopez we can test the null hypothesis that the $QPS(\overline{GDP}_{01}, GDP_{01})$ is the same as the $QPS(f, GDP_{01})$, in other words under the null hypothesis the skill score in equation 11 is equal to zero. If $\bar{d}$ represents the average of:

$$d_t = 2(GDP_{01t} - f_t)^2 - 2(GDP_{01t} - 0.31)^2 \tag{12}$$

and $f_d(0)$ is the spectral density function at frequency zero, then asymptotically the Lopez test statistic $S_1 = \frac{\bar{d}}{\sqrt{2\pi f_d(0)/T}}$ is ~ $N(0, 1)$ under the null hypothesis.

The QPS and SS statistics are calculated for each of the proposed aggregation methods and reported in Table 4.3 together with the correlation, AUC-ROC and p-value of the Lopez test.

Table 4.3 – Assessing the performance of the aggregation methods

Statistics	Probability Prediction				Binary Prediction			
	Amean	Gmean	Elop	Beta	Bench	Youden	Chi.sq	CTL
				Hybrid approach				
Correlation	0.66	0.66	0.66	0.69	0.74	0.74	0.73	0.77
AUC-ROC	0.93	0.94	0.93	0.93	0.97	0.97	0.96	0.98
QPS	0.10	0.10	0.09	0.07	0.09	0.08	0.07	0.07
SS	0.19	0.26	0.27	0.47	0.31	0.37	0.46	0.46
Lopez Test *p*-value	0.22	0.09	0.07	0.00	0.08	0.03	0.00	0.01
				AUC-ROC selection				
Correlation	0.63	0.62	0.62	0.59	0.66	0.67	0.66	0.69
AUC-ROC	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.95
QPS	0.10	0.02	0.02	0.09	0.09	0.09	0.08	0.08
SS	0.26	0.29	0.29	0.30	0.28	0.30	0.36	0.37
Lopez Test *p*-value	0.12	0.09	0.08	0.09	0.18	0.13	0.05	0.06
				Coherence selection				
Correlation	0.41	0.44	0.42	0.48	0.41	0.47	0.49	0.49
AUC-ROC	0.72	0.72	0.72	0.72	0.71	0.77	0.77	0.76
QPS	0.14	0.13	0.13	0.10	0.15	0.13	0.11	0.12
SS	- 0.11	- 0.05	- 0.04	0.22	- 0.19	- 0.05	0.13	0.07
Lopez Test *p*-value	0.43	0.73	0.77	0.03	0.23	0.77	0.36	0.66

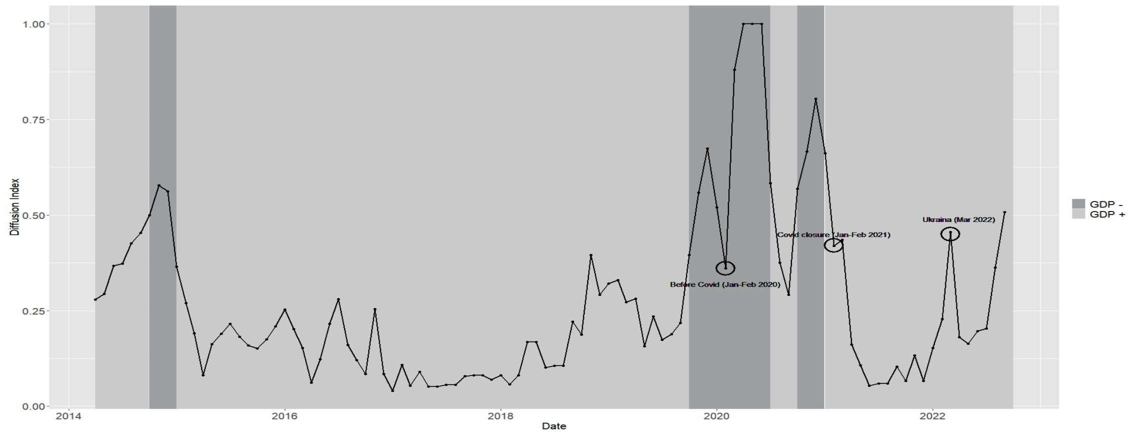
Source: Authors' elaboration

In order to assess the performance of the selection methods, the statistics are calculated for: AUC-ROC selection (that uses only the AUC-ROC criterium to select the series), Coherence selection (based only on the spectral coherence criterium), and the our hybrid approach which exploits both AUC-ROC and spectral coherence criteria. The first result that stands out from the Table 4.3, is the better performance of the hybrid approach compared to the other two methods, confirming the idea that the mix of criteria improves nowcasting capabilities.

Overall the statistics for the binary point prediction approaches are better than the ones of the probability prediction approaches. Between the latter the beta approach seems to work better in terms of SS and Lopez test p-value. However the results suggest that the closest topleft method (CTL) shows the highest value for the AUC-ROC and for the correlation with GDP_{01} having in terms of QPS a good performance as well.Looking at the p-value of the Lopez test, the CTL together with the Chi sq aggregation method exhibits significant skill compared with constant base rate forecast.

On the basis of all the evidence reported, the CTL indicator, plotted in Figure 4.7, is the method that gives the best results and will be our proposed Rat-ita indicator. In particular we would like to stress the reaction of our indicator to exogenous shocks. The three points highlighted with a black circle correspond to events that occurred in the most recent years. As a consequence our indicator gives different signal within the quarter compared to the sign of the final quarterly GDP growth rate (reported in the figure as grey area for negative sign and white area for positive sign). For example the mixed signal emerging from the indicator value for March 2022 compared to the GDP growth rate was driven by the turmoil in the economic system due to the start of the war in Ukraine.

Figure 4.7 – Best diffusion index



Source: Authors' elaboration

4.4 RAT-Ita performance against benchmark models

In order to evaluate RAT-Ita forecast performance, its nowcast have been compared with those of traditional benchmark models. In particular we used three different models to forecast the q-o-q GDP growth rate: an ARIMA model, a bridge model exploiting the quarterly averages of m-o-m industrial production index growth rate (see *e.g.* Golinelli and Parigi, 2007), and a mixed frequency model using the monthly m-o-m industrial production index growth rate (see *e.g.* Ghysels *et al*., 2007).

Given that RAT-Ita nowcasts the sign of the q-o-q GDP growth rate, the GDP growth rates predicted by the benchmark models over the period Q2-2014 to Q2-2022 are transformed in dummy variables equal to 0 if the growth rate was positive and to 1 if negative. The statistics used for the comparison are:

- Accuracy: the proportion of correctly classified cases out of the total number of cases;
- Precision: the proportion of negative growth rates correctly ranked out of the number of times the growth rate was predicted to be negative;
- Specificity: the proportion of correctly classified positive growth rates out of the total positive growth rates;
- Sensitivity: the proportion of correctly classified negative growth rates out of the total negative growth rates.

In addition, we also computed both DAC p-values and AUC-ROC statistics. Concerning all the indicators, RAT-Ita shows a better performance in the forecast the sign of GDP q-o-q growth rates (Table 4.4) compared to the other models. The Midas model is ranked as the second best, confirming on the importance on the availability of the monthly information on IPI for the forecast of GDP.

Table 4.4 – The Performance of benchmark models and RAT-Ita

	ARIMA	Bridge	Midas	RAT-Ita
Accuracy	0.45	0.55	0.51	**0.96**
Precision	0.16	0.16	0.20	**0.87**
Specificity	0.43	0.56	0.48	**0.98**
Sensitivity	0.6	0.47	0.73	**0.87**
DAC *p*-value	0.97	0.74	0.38	**0.00**
AUC	0.7	0.44	0.59	**0.98**

Source: Authors' elaboration

4.5 RAT-Ita over 2022: implication for dissemination

In this section we would like to illustrate the implication for the dissemination of RAT-Ita as a new tool to monitor the evolution of the Italian business cycle.

For each quarter, RAT-Ita estimates three times (months), the sign of qo-q GDP growth. Consequently, three different scenarios are available for dissemination:

- if all the three estimates of the sign are near to 1, we predict a negative GDP growth for the quarter;
- if all the three estimates of the sign are near to 0, we predict a positive GDP growth for the quarter;
- if the three estimates of the sign result as a combination of 0 to 1, uncertainty emerges about the sign of the GDP growth.

To illustrate this point we consider the first 9 months of 2022. Table 4.5 reports the shares of the selected indicators predicting negative signs for each block together with the monthly values of Rat-Ita. In the second part of the table we present the q-o-q GDP growth.

For example considering Q1, the selected indicators belonging to the domain of index of industrial production showed a negative performance in the three months (more than the 50% were negative) reinforced, in March, by a high negative share of the business climate in services. Even the share of the selected time series for the Business climate in manufacturing surveys were deteriorating in the same month. The estimated values of RAT-Ita for January, February and March returns a suggestion for a positive q-o-q of GDP growth but with a deteriorating of the economic activity in March.

In Q2 all the selected indicators showed mainly a positive behaviour with the exception of the IPI in May, where 50% experimented a negative sign. The very low values of RAT-Ita for all the months suggest for a positive q-o-q GDP growth rate.

Considering Q3, the value for RAT-ta were 0.21, 0.36 and 0.51 respectively for July, August and September showing a deterioration of the evolution of the GDP across the quarter. Looking at the domain, the negative evolution is mainly driven by the selected indicators of the index of industrial production and of the business climate in manifacturing where the share of indicators with negative sign comes up along the quarter, from 33% and 23% in July to 67% and 51% in September. The evolution of the business climate in services follows a different pattern, characterised by a low level of the share of the indicators with a negative signal.

These evidences are in lines with the results of the value added in the third quarter that showed a negative q-o-q growth rate for manufacturing and construction and a positive one for services.

The results for Q3 provide also a clear evidence of the present limits in the use of RAT-Ita in the real-time: the behind methodology is able to track for the evolution of the economic activity but does not return yet a mapping among the value of the monthly indicator and the expected value for the GDP growth rate. Neverthless, we are able to argue that having two out of three months related to a positive sign for the GDP growth rate, together with a not generalised deteriorating phase across the sectors oh the third month, might support the hypothesis of the persisting of the expansion of the economic activity.

Table 4.5 – Percentage of the selected indicators with negative sign by block. Year 2022 (a)

Domain	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
External Trade	-	-	-	100	0	0	0	100	100
Labour Market (b)	0	0	33	0	0	0	0	33	50
Business climate in manufacturing	4	10	40	20	17	24	23	35	51
Index of Industrial Production	50	70	50	33	50	33	33	67	67
Business climate in services	36	43	64	0	0	0	15	23	38
RAT-Ita	**0.15**	**0.23**	**0.46**	**0.18**	**0.16**	**0.2**	**0.21**	**0.36**	**0.51**

Source: Authors' elaboration
(a) q-o-q GDP growth: Q1 = +0.1%; Q2 = +1.1%; Q3 = +0.5%.
(b) Only 1 series has been selected for this domain.

Overall, to our knowledge, both the evidences on the negative contribution of the selected indicators and the the value of the monthly RAT-Ita improves the short-term narrative for the Italian business cycle.

5. Concluding remarks

Monitoring the business cycle evolution during challenging times is a difficult job that asks for timely information, exploration of new data sources, and new methodologies to better exploit them. This paper presents a novel approach to exploit a large information set of 1,285 monthly indicators to target the sign of one-quarter-ahead GDP growth (*i.e.* the direction of GDP changes).

To do so, we introduce a tool which selects in each quarter a subset of indicators able to detect GDP ups and downs, and aggregates the forecasts of the single series. The selection step of our approach jointly relies on the Directional Accuracy Changes, on the Receiver Operating Characteristic, and on the spectral coherence of indicators with GDP growth over the cyclical frequencies (2-8 years). The aggregation step relies on bi-variate logit models and on the use of binary point prediction based on the ROC.

The proposed methodology has been tested by analysing the Italian case, looking at its performance both in pseudo real-time assessments of the predictive ability against several benchmark models, and at its ability to explain the evolution of the business cycle over the first nine months of 2022.

Although its very promising performance, RAT-Ita may be further improved in both the theoretical and the empirical sides. Firstly, more effort could be paid to set better theoretically-founded thresholds, which this work has simply explored on the empirical ground. Secondly, confidence intervals of RAT-Ita predictions could be derived in order to deepen their information. Thirdly, the present investigation could be extended to other countries, such as the the Euro area, or be applied to more specific targets, such as consumption, investment, and external trade variables.

Finally, the use of RAT-Ita can be combined, at least for the Italian case, with the Macroeconometric Model MeMo-It (Bacchini *et al.* (2013), Bacchini *et al.* (2018), which is based on annual data but that can embody short-term predictions from external (high-frequency) sources of information.

References

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese. 2010. “New Eurocoin: Tracking Economic Growth in Real Time”. *The Review of Economics and Statistics*, Volume 92, N. 4: 1024-1034.

Altissimo, F., D.J. Marchetti, and G.P. Oneto. 2000. “The Italian business cycle: Coincident and Leading Indicators and Some Stylized Facts”. *Giornale degli Economisti e Annali di Economia*, Nuova serie, Volume 60 (Anno 113), N. 2: 147-220.

Aprigliano, V., S. Emiliozzi, G. Guaitoli, A. Luciani, J. Marcucci, and L. Monteforte. 2023. “The power of text-based indicators in forecasting Italian economic activity”. *International Journal of Forecasting*, Volume 39, Issue 2: 791-808.

Aruoba, S.B., F.X. Diebold, and C. Scotti. 2009. “Real-time Measurement of Business Conditions”. *Journal of Business & Economic Statistics*, Volume 27, Issue 4: 417–427.

Babii, A., E. Ghysels, and J. Striaukas. 2022. “Machine Learning Time Series Regressions With an Application to Nowcasting”. *Journal of Business & Economic Statistics*, Volume 40, Issue 3: 1094-1106.

Bacchini, F., M.E. Bontempi, R. Golinelli, and C.S. Jona-Lasinio. 2018. “Short- and long-run heterogeneous investment dynamics”. *Empirical Economics*, Volume 54, Issue 2: 343–378.

Bacchini, F., C. Brandimarte, P. Crivelli, R. De Santis, M. Fioramanti, A. Girardi, R. Golinelli, C.S. Jona-Lasinio, M. Mancini, C. Pappalardo, D. Rossi, M. Ventura, and C. Vicarelli. 2013. “Building the core of the Istat system of models for forecasting the Italian economy: MeMo-It”. *Rivista di statistica ufficiale/Review of official statistics*, N. 1/2013: 17–45. Roma, Italy: Istat. https://www.istat.it/it/archivio/101680.

Baffigi, A., R. Golinelli, and G. Parigi. 2004. “Bridge models to forecast the euro area GDP”. *International Journal of Forecasting*, Volume 20, Issue 3: 447-460.

Bai, J., and S. Ng 2002. “Determining the Number of Factors in Approximate Factor Models”. *Econometrica*, Volume 70, Issue 1: 191–221.

Baker, S.R., N. Bloom, and S.J. Davis. 2016. “Measuring Economic Policy Uncertainty”. *The Quarterly Journal of Economics*, Volume 131, Issue 4: 1593-1636.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin. 2013. “Now-Casting and the Real-Time Data Flow”. In Elliott, G., and A. Timmermann *(Eds.). Handbook of Economic Forecasting*, Volume 2, Part A, Chapter 4: 195-237. Amsterdam, The Netherlands: Elsevier.

Barbaglia, L., L. Frattarolo, L. Onorante, F.M. Pericoli, M. Ratto, and L. Tiozzo Pezzoli. 2022. “Testing big data in a big crisis: Nowcasting under COVID-19”. *International Journal of Forecasting*: Article in press.

Berge, T.J., and Ò. Jordà. 2011. “Evaluating the Classification of Economic Activity into Recessions and Expansions”. *American Economic Journal: Macroeconomics*, Volume 3, N. 2: 246-277.

Boivin, J., and S. Ng. 2006. "Are more data always better for factor analysis?". *Journal of Econometrics*, Volume 132, Issue 1: 169-194.

Bry, G., and C. Boschan. 1971. "Front matter to 'Cyclical Analysis of Time Series: Selected Procedures and Computer Programs'. In Bry, G., and C. Boschan. "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs". National Bureau of Economic Research - NBER *Technical Paper*, N. 20. Cambridge, MA, U.S.: NBER.

Bulligan, G., R. Golinelli, and G. Parigi. 2010. "Forecasting monthly industrial production in real-time: from single equations to factor-based models". *Empirical Economics*, Volume 39, Issue 2: 303-336.

Chow, G.C., and An-loh Lin. 1971. "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series". *The Review of Economics and Statistics*, Volume 53, N. 3: 372-375.

Costa, S., F. Sallusti, C. Vicarelli, and D. Zurlo. 2019. "Over the ROC methodology: Productivity, economic size and firms' export thresholds". *Review of International Economics*, Volume 27, Issue 3: 955-980.

Coulombe, P.G., M. Marcellino, and D. Stevanović. 2021. "Can machine learning catch the COVID-19 recession?". *National Institute Economic Review - NIER*, Volume 256: 71-109.

Cubadda, G., B. Guardabascio, and A. Hecq. 2013. "A general to specific approach for constructing composite business cycle indicators". *Economic Modelling*, Volume 33: 367-374.

Diebold, F.X., and R.S. Mariano. 2002. "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics*, Volume 20, Issue 1: 134-144.

Estrella, A., and G.A. Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity". *The Journal of Finance*, Volume 46, N. 2: 555-576.

Fawcett, T. 2006. "An introduction to ROC analysis". *Pattern Recognition Letters*, Volume 27, Issue 8: 861-874.

Fernández, R.B. 1981. "A Methodological Note on the Estimation of Time Series". *The Review of Economics and Statistics*, Volume 63, N. 3: 471–476.

Forni, M., A. Giovannelli, M. Lippi, and S. Soccorsi. 2018. "Dynamic factor model with infinite-dimensional factor space: Forecasting". *Journal of Applied Econometrics*, Volume 33, Issue 5: 625-642.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin. 2001. "Coincident and Leading Indicators for the Euro Area". *The Economic Journal*, Volume 111, N. 471 (*Conference Papers*): C62–C85.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin. 2000. "The Generalized Dynamic-Factor Model: Identification and Estimation". *The Review of Economics and Statistics*, Volume 82, N. 4: 540-554.

Fortin-Gagnon, O., M. Leroux, D. Stevanović, and S. Surprenant. 2022. "A Large Canadian Database for Macroeconomic Analysis". *Canadian Journal of Economics/Revue Canadienne d'Économique*, Volume 55, Issue 4: 1799-1833.

Ghysels, E., A. Sinko, and R. Valkanov. 2007. "MIDAS Regressions: Further Results and New Directions". *Econometric Reviews*, Volume 26, Issue 1: 53-90.

Golinelli, R., and G. Parigi. 2007. "The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries". *Journal of Forecasting*, Volume 26, Issue 2: 77-94.

Gómez, V., and A. Maravall. 1996. "Programs TRAMO and SEATS. Instructions for the user (beta version: September 1996)". *Documento de Trabajo*, N. 9628. Madrid, Spain: Banco de España, Servicio de Estudios.

Harding, D., and A. Pagan. 2002. "Dissecting the cycle: a methodological investigation". *Journal of Monetary Economics*, Volume 49, Issue 2: 365–381.

Kim, K., H.H. Kim, and N.R. Swanson. 2023. "Mixing mixed frequency and diffusion indices in good times and in bad: an assessment based on historical data around the great recession of 2008". *Empirical Economics*, Volume 64, Issue 3: 1421-1469.

Kuzin, V., M. Marcellino, and C. Schumacher. 2011. "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area". *International Journal of Forecasting*, Volume 27, Issue 2: 529-542.

Lahiri, K., and J.G. Wang. 2013. "Evaluating probability forecasts for GDP declines using alternative methodologies". *International Journal of Forecasting*, Volume 29, Issue 1: 175–190.

Lahiri, K., and J.G. Wang. 2006. "Subjective Probability Forecasts for Recessions". *Business Economics*, Volume 41, Issue 2: 26-37.

Lahiri, K., and L. Yang. 2018. "Confidence Bands for ROC Curves With Serially Dependent Data". *Journal of Business & Economic Statistics*, Volume 36, Issue 1: 115-130.

Lahiri, K., and L. Yang. 2013. "Forecasting Binary Outcomes". In Elliott, G., and A. Timmermann (Eds.). *Handbook of Economic Forecasting*, Volume 2, Part B, Chapter 19: 1025-1106. Amsterdam, The Netherlands: Elsevier.

Lewis, D.J., K. Mertens, J.H. Stock, and M. Trivedi. 2022. "Measuring real activity using a weekly economic index". *Journal of Applied Econometrics*, Volume 37, Issue 4: 667-687.

Litterman, R.B. 1983. "A Random Walk, Markov Model for the Distribution of Time Series". *Journal of Business & Economic Statistics*, Volume 1, Issue 2: 169-173.

Lopez, J.A. 2001. "Evaluating the predictive accuracy of volatility models". *Journal of Forecasting*, Volume 20, Issue 2: 87-109.

Lusted, L.B. 1960. "Logical analysis in roentgen diagnosis. Memorial fund lecture". *Radiology*, Volume 74, N. 2: 178-193.

McCracken, M.W., and S. Ng. 2020. "FRED-QD: A Quarterly Database for Macroeconomic Research". *Working Paper*, N. 2020-005B. St. Louis, MO, U.S.: Federal Reserve Bank of St. Louis, Economic Research.

McCracken, M.W., and S. Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research". *Journal of Business & Economic Statistics*, Volume 34, Issue 4 (Special Issue on Big Data): 574-589.

Miller, R., and D. Siegmund. 1982. "Maximally Selected Chi Square Statistics". *Biometrics*, Volume 38, N. 4: 1011-1016.

Mitchell, W.C., and A.F. Burns (Eds.). 1938. *Statistical indicators of cyclical revivals*. Cambridge, MA, U.S.: National Bureau of Economic Research - NBER.

Pesaran, M.H., and A. Timmermann. 1992. "A Simple Nonparametric Test of Predictive Performance". *Journal of Business & Economic Statistics*, Volume 10, Issue 4: 461-465.

Proietti, T., A. Giovannelli, O. Ricchi, A. Citton, C. Tegami, and C. Tinti. 2021. "Nowcasting GDP and its components in a data-rich environment: The merits of the indirect approach". *International Journal of Forecasting*, Volume 37, Issue 4: 1376–1398.

Ranjan, R., and T. Gneiting. 2010. "Combining Probability Forecasts". *Journal of the Royal Statistical Society*, Series B (Statistical Methodology), Volume 72, N. 1: 71-91.

Satopää, V.A., J. Baron, D.P. Foster, B.A. Mellers, P.E. Tetlock, and L.H. Ungar. 2014. "Combining multiple probability predictions using a simple logit model". *International Journal of Forecasting*, Volume 30, Issue 2: 344-356.

Stock, J.H., and M.W. Watson 2002a. "Forecasting Using Principal Components from a Large Number of Predictors". *Journal of the American Statistical Society*, Volume 97, N. 460: 1167-1179.

Stock, J.H., and M.W. Watson. 2002b. "Macroeconomic Forecasting Using Diffusion Indexes". *Journal of Business & Economic Statistics*, Volume 20, Issue 2: 147-162.

Stock, J.H., and M.W. Watson. 1992. "A Procedure for Predicting Recessions With Leading Indicators: Econometric Issues and Recent Experience". National Bureau of Economic Research - NBER *Working Paper*, N. 4014. Cambridge, MA, U.S.: NBER.

Stock, J.H., and M.W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators". In Blanchard, O.J., and S. Fischer *(Eds.)*. *NBER Macroeconomics Annual 1989*, Volume 4: 351–394. Cambridge, MA, U.S.: MIT Press.

Yang, L., K. Lahiri, and A. Pagan. 2023. "Getting the ROC into Sync". *Journal of Business & Economic Statistics*, Published online: 3rd January 2023. https://doi.org/10.1080/07350015.2022.2154778.

Youden, W.J. 1950. "Index for rating diagnostic tests". *Cancer*, Volume 3, Issue 1: 32-35.

Informazioni per le autrici e per gli autori

La collana è aperta alle autrici e agli autori dell'Istat e del Sistema statistico nazionale e ad altri studiosi che abbiano partecipato ad attività promosse dall'Istat, dal Sistan, da altri Enti di ricerca e dalle Università (convegni, seminari, gruppi di lavoro, etc.).

Coloro che desiderano pubblicare su questa collana devono sottoporre il proprio contributo al Comitato di redazione degli Istat working papers, inviandolo per posta elettronica all'indirizzo: iwp@istat.it.

Il saggio deve essere redatto seguendo gli standard editoriali previsti (disponibili sul sito dell'Istat), corredato di un sommario in Italiano e in Inglese e accompagnato da una dichiarazione di paternità dell'opera.

Per le autrici e gli autori dell'Istat, la sottomissione dei lavori deve essere accompagnata da un'e-mail della/del propria/o referente (Direttora/e, Responsabile di Servizio, etc.), che ne assicura la presa visione.

Per le autrici e gli autori degli altri Enti del Sistan la trasmissione avviene attraverso la/il responsabile dell'Ufficio di statistica, che ne prende visione. Per tutte le altre autrici e gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione.

Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del Chicago Manual of Style.

Attraverso il Comitato di redazione, tutti i lavori saranno sottoposti a un processo di valutazione doppio e anonimo che determinerà la significatività del lavoro per il progresso dell'attività statistica istituzionale.

La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line gratuitamente.

Gli articoli pubblicati impegnano esclusivamente le autrici e gli autori e le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.