



GDP Forecasting: Machine Learning, Linear or Autoregression?

Giovanni Maccarrone¹, Giacomo Morelli^{2*} and Sara Spadaccini^{3,4}

¹Department of Economic and Social Sciences, Sapienza University of Rome, Rome, Italy, ²Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy, ³Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Rome, Italy, ⁴Global Data Hub, Enel Global Services S.r.l., Rome, Italy

This paper compares the predictive power of different models to forecast the real U.S. GDP. Using quarterly data from 1976 to 2020, we find that the machine learning K-Nearest Neighbour (KNN) model captures the self-predictive ability of the U.S. GDP and performs better than traditional time series analysis. We explore the inclusion of predictors such as the yield curve, its latent factors, and a set of macroeconomic variables in order to increase the level of forecasting accuracy. The predictions result to be improved only when considering long forecast horizons. The use of machine learning algorithm provides additional guidance for data-driven decision making.

Keywords: k nearest neighborhood, machine learning, time series, GDP, forecasting strategies

OPEN ACCESS

Edited by:

Emanuela Raffinetti,
University of Pavia, Italy

Reviewed by:

Thomas Edward Leach,
University of Pavia, Italy
Paolo Pagnottoni,
University of Pavia, Italy

*Correspondence:

Giacomo Morelli
giacomo.morelli@uniroma1.it

Specialty section:

This article was submitted to
Artificial Intelligence in Finance,
a section of the journal
Frontiers in Artificial Intelligence

Received: 12 August 2021

Accepted: 20 September 2021

Published: 15 October 2021

Citation:

Maccarrone G, Morelli G and
Spadaccini S (2021) GDP Forecasting:
Machine Learning, Linear or
Autoregression?
Front. Artif. Intell. 4:757864.
doi: 10.3389/frai.2021.757864

1 INTRODUCTION

The real Gross Domestic Product (GDP) is a single, omni-comprehensive measure of the economic activity that considers the total value of goods and services produced in the economy. It is considered by academics, investors, and regulators as a proxy for the wealth of the economy and an informative indicator that drives the decision-making processes (Provost and Fawcett, 2013). This makes the forecast of the GDP a relevant issue. Indeed, it is of interest to target national economic policies as well as in other fields, from non-performing loans (Bouheni et al., 2021) to natural disaster (Atsalakis et al., 2020).

When the research question is the forecast of periods of growth or recession a popular methodology is to decompose the GDP in cyclical and trend components relying on appropriate filters. A growth (recession) means that the value of the cycle component is positive (negative) for a given period. However, within this approach it is only possible to assess the growth (or recession) losing the quantitative information on the prediction. This is instead achieved through the regression methodology. That it is the approach we follow, such choice being driven by the limitations encountered in the decomposition of the GDP in cyclical and trend components (Luginbuhl and Koopman, 2004).

The aim of this paper is to show among a set of different models and forecasting strategies which performs better. Classical time series analysis or machine learning? One-step-ahead or multi-step-ahead forecast? Including macro-economic variables or just the self-explanatory GDP values? How does our model respond to periods of economic turbulence? These are the research questions we aim to provide an answer.

Several approaches have been proposed in the literature to forecast the GDP. Indeed, the macroeconomic literature that investigates this topic through the time series approach mainly use different specifications of VAR (Ang et al., 2006; Brave et al., 2019; Koop et al., 2020), and forecasting improvements can be achieved relying on appropriate Bayesian shrinkage procedures, as highlighted in Bańbura et al. (2010). Regarding the potential economic indicators that are used as

predictors of the GDP, many authors converge on the use of the yield curve that contains information about future economic activity (Giannone et al., 2008; Yiu and Chow, 2010). Estrella and Hardouvelis (1991) find that especially the slope of the yield curve can predict cumulative variations in real GDP for up to 4 years into the future. A similar study is carried out by Bernard and Gerlach (1998) in eight countries finding that, although there are substantial differences across the countries, the slope of the yield provides information about the possibility of future recessions, whereas Ang et al. (2006) find that nominal short rates outperform the slope of the yield curve in forecasting GDP growth. Other studies (Koop, 2013; Schorfheide and Song, 2015) use instead a set of macroeconomic variables to predict the U.S. GDP. Drawing from this strand of literature, e.g., Estrella and Mishkin (1996), Koop (2013), Diebold et al. (2006), we use the yield curve as well as its latent factors and a set of macroeconomic variables, namely Consumer Price Index, Unemployment rate, Federal Fund rates, and Manufacturing Capacity Utilization.

Chauvet and Potter (2013) offer a comparison between reduced form, autoregressive, VAR, and Markov switching models and find that simple time series autoregressive process of order two [AR (2)] outperforms other models in the forecast of the U.S. GDP. Baffigi et al. (2002) provide an example of the use of ARIMA for the U.E. GDP prediction. Lunde and Torkar (2020) exploit more than 120 predictors and then perform a principal component analysis (PCA) to reduce the number of variables. Despite the inclusion of different sources of information in their set-up, the PCA does not provide the economic interpretation of the results.

In this paper, we propose models with macro-economic variables and other models that take advantage of the self-explanatory information of the GDP relying on both classical time series analysis as well as on a machine learning algorithm. In particular, we forecast the U.S. GDP with ARX, SARIMAX and Linear Regression to include additional information such as real and financial measures of economic activity, and use AR and SARIMA as a benchmark for time series analysis. We also exploit the K-Nearest Neighbour (KNN) machine learning methodology. Our goal is to achieve forecasts with high accuracy and with high degree of explainability that is a best practice for building trust between machine learning and decision-makers, as pointed out in Bellotti et al. (2021). The idea is that the decision-maker should adopt the machine learning as a powerful instrument and should employ it with awareness without regarding it as a “black-box.”

Many studies explore the potential of machine learning in the field of forecasting. Stone (1977) shows the consistency property of the non-parametric KNN estimator. The model is widely used for classification tasks such as object identification and, due to the easy implementation and explainability, it is also used in applications such as missing data imputation (Bertsimas et al., 2021) and reduction of training set (Wauters and Vanhoucke, 2017) being able to better identify similar objects. The KNN can identify repeated patterns within the time series and for this reason is applied to financial time series modeling as in Ban et al. (2013). Al-Qahtani and Crone (2013) use KNN for forecasting U.K. electricity demand and find that KNN outperforms better forecasts than other benchmark models. Rodríguez-Vargas

(2020) finds that KNN outperforms also two competitors machine learning models, the random forest and the extreme gradient boosting, in terms of accuracy for predicting the inflation. In general, KNN has been referenced as one of the top ten algorithms in data mining (Wu et al., 2008). Moreover, KNN is especially suitable for cases in which there is not an high number of past observations, i.e., very little past information. As pointed out in Wauters and Vanhoucke (2017), artificial intelligence methods require a minimum number of observation to work properly whereas for the KNN this limitation is not so strict even though a minimum number of observation is required (Diebold and Nason, 1990). We therefore employ the KNN model as it offers a simple methodology based on distance metrics to exploit past information.

A compelling way to predict real economic variables is offered by the nowcasting literature, which aims to predict their values in the very short term. When the objective is to study the prediction at horizons lower than a quarter, given quarterly data available for GDP, it is possible to use a consistent two step estimator, as in Doz et al. (2011), that provides the policymaker with an early estimate of the next quarter including auxiliary exogenous predictors available at a lower frequency. Moreover, this framework can be empowered with alternative variables to boost the economic knowledge. For instance, Spelta and Pagnottoni (2021) use nowcasting to assess the impact of mobility restrictions on the economic activity during the pandemic. In particular, they study the trade-off between economic sacrifices and health outcomes in terms of timely policy suggestions. Foroni et al. (2020) explicitly focus on the forecast and nowcast of COVID-19 recession and recovery studying the GDP growth and showing an interesting similarity with the great recession.

We analyze two different forecasting strategies: the one-step-ahead and the multi-step-ahead forecasts (Marcellino et al., 2006; Hu et al., 2020). The former is more reliable and accurate by construction, however it results to be less informative for macroprudential policies. In the multi-step-ahead strategy proposed, we forecast the U.S. GDP up to 12 quarters in advance. This information is potentially extremely valuable although much more challenging.

Finally, we evaluate the performance in terms of mean square error. In particular, we are interested in studying the trade-off between two different aspects: the accuracy of the estimates even when considering a period of economic turbulence, and the forecasting horizon.

The rest of the paper is organized as follows: **Section 2** introduces the model specifications and the empirical strategy, **Section 3** illustrates the empirical analysis, **Section 4** reports the results and **Section 5** concludes.

2 MODEL SPECIFICATIONS AND EMPIRICAL STRATEGY

2.1 Motivations

Closely related to the GDP forecast is the ability to understand whether the forecasted value is associated with growth or

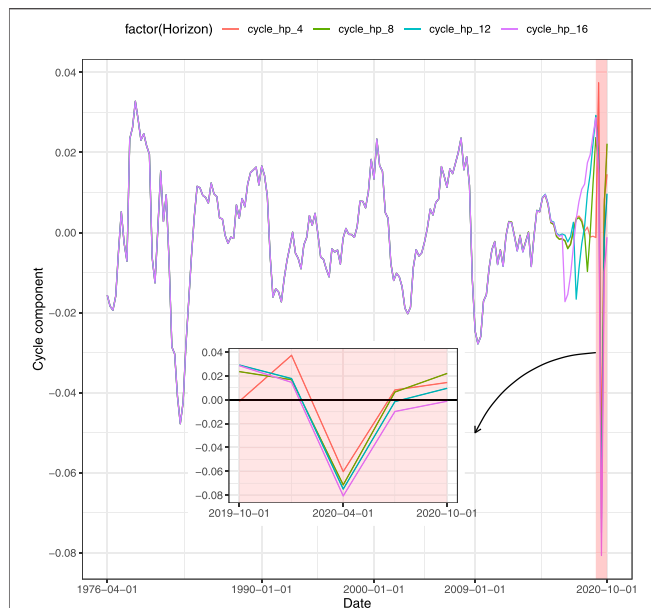


FIGURE 1 | Cycle decomposition of GDP, sensitiveness to different horizons until last quarter of 2020.

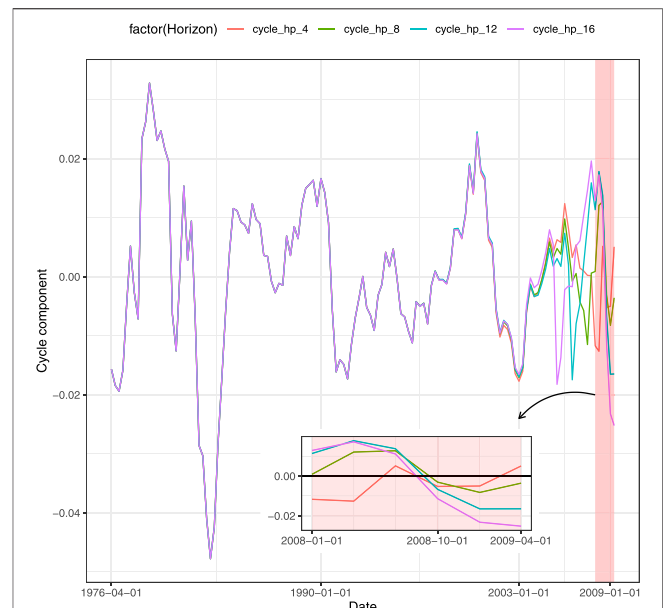


FIGURE 2 | Cycle decomposition of GDP, sensitiveness to different horizons until second quarter of 2009.

recession for the economy. It can be achieved through a classification framework that defines a binary target variable starting from the time series of the GDP. An appealing approach to detect recessions is to decompose the GDP in trend and cyclical components. Among the techniques used, the filters are the most employed in literature. A well-known technique is the Hodrick and Prescott (1997) filter, also known as H-P filter, which through an appropriate parametrized minimization problem generates the GDP cycle component. Once the cycle component has been detected from the time series, it is then transformed into a binary variable that assumes value equal to 1 (recession) whenever the cyclical component is lower than zero and 0 (growth) otherwise. Nevertheless, the use of this approach has been criticized. Hamilton (2018) proposes a regression filter as an alternative. Even if such regression filter overcomes the drawbacks of the H-P filter, it results to suffer some limitations, as discussed in Schüler (2018). Another procedure as in Bernard and Gerlach (1998) and Estrella and Hardouvelis (1991) is to set the GDP equal to unit during the quarters of recession indicated by the National Bureau of Economic Research (NBER).¹ Applying the H-P filter to our data, we have encountered the limitations of this filter on the right tail. In **Figure 1** each line represents a different size of the test set when splitting the entire time series into train and test sets. The red line shows the value of the cycle when the test set contains the 4 quarters of 2020, the green line does the job for 2 years (8 quarters) and so on. The feature that clearly emerges is that the values obtained through the filter are affected by the size of the test set. Using the test set with the last 4 and 8 quarters, the H-P filter assigns to the third quarter of 2020 a positive value. This

means that the classification procedure on the filter generates those quarters as periods of growth (rather than recessions). As a result, the policymaker waste resources since the model is being fitted on unreliable data. When the test size is long enough, the filter provides the policymaker with appropriate values. Notice that the value obtained comparing the binary outcome derived from the H-P filter and the NBER data, that is the one for which the two time series match is 12 quarters in our example. We also control for the Subprime recession. Similarly, more than 4 quarters are required by H-P filter to match the NBER recession period for the second quarter of 2009, as shown in **Figure 2**. Since H-P filter cannot be considered reliable on the tails, the classification approach does not represent a trustworthy model for predicting growth (recession). Furthermore, another drawback of the classification is the loss of the quantitative information: the decision maker is provided with signal of growth or recession without any kind of information related to the magnitude of the event. We point out that neglecting such quantitative specification comes at a cost as the resulting classification will rely on biased trend-cycle decomposition and, therefore, be misleading. Instead, using predictions based on the actual value of the GDP, the benefit for the policymaker is to capture the intensity of the variation. In this way, the entity of the growth (recession) of the GDP assumes a real value that can be fundamental to address medium-term economic policies. In contrast to the cyclical indicator, this type of information gives the policymaker a wider set of possible actions than a binary pair (growth or recession), to better calibrate the reaction to expected changes in the GDP. For instance, the Federal Reserve System (FRS) may be interested in the GDP growth forecast with the aim to set the interest rate against any inflationary threats. On the one hand, when the forecast is based on classification, the only strategy the FRS can apply is to lower or raise the interest rate

¹NBER considers as recessions two consecutive quarters of negative GDP growth

without knowledge of the value which is needed to set the policy. On the other hand, a quantitative information about the prediction of the GDP growth allows the FRS to optimally set the interest rate, following classical policies such as the Taylor rule (Taylor, 1993) or other rules, to respond to variation of the GDP. For all these reasons, we forecast the GDP with regression techniques.

2.2 GDP Forecasting Models

We explore different forecasting models to predict the United States GDP: KNN, AR, SARIMA, ARX, SARIMAX, and a particular specification of the classical linear regression model (LR). Let $t_i, i \in \{1, 2, 3, 4\}$ represents the t_i -th quarter of year $T \in \{1976, 1977, \dots, 2020\}$, so that $t \in \{1, \dots, 179\}$ is the number of total quarters. Define $Y = \{y_t\}_{t \in \mathbb{R}^+}$ the time series of the log GDP. Let $Y^d = \{y_t - y_{t-d}\}_{d, t \in \mathbb{R}^+}$ be the d -th order difference between consecutive GDP time series observations. We denote with $\mathbf{X}_n = \{\mathbf{x}_{n,t}\}$ the time series of a generic set of n covariates with $n \in \{1, 2, 3, \dots, N\}$.

KNN. The KNN is a machine learning algorithm useful to solve both classification and regression problems (Wu et al., 2008) based on learning by analogy. We apply the KNN methodology to forecast univariate time series. The rationale behind the use of KNN for time series forecasting is that a time series may contain repetitive patterns. The i -th data point (target) can be described by a vector of n covariates $(x_1^i, x_2^i, \dots, x_N^i)$ that are the lagged values of the target y_i^1 . Consider a new observation, for example the next quarter y_{t+1}^1 to be predicted, whose covariates are known and denoted as $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$. Note that there is a relationship between the covariates of the new observations that we want to forecast and the information that we have. The last targets are used as covariates of the new observation. Given that the minimum lag must be at least equal to the number of periods of forecast, in our analysis we use one covariate. For example, if the forecasting period is $h = 10$, the target y_{t+1}^1 will be described by the covariate x_{t-10} . The KNN algorithm exploits the covariates of the new observation to find the k most similar training covariates according to a specified distance metric. In this study, we use as similarity metric the euclidean distance between the new observation $t + 1$ and the i -th training observation:

$$\sqrt{\sum_{n=1}^N (x_n^i - \bar{x}_n)^2}. \tag{1}$$

When predicting a new data point, the algorithm finds the k observed targets with covariates' values (the x lagged quarters) closer to it. Then, it assigns to the new data point the average of the k 's target values. We use *tsfknn* library on the software *R* for the implementation (Martínez et al., 2019).

AR. The purely autoregressive process of order p , $AR(p)$ satisfies the equation:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \tag{2}$$

where $\{\varepsilon_t\}$ is a white noise with $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2$, p is the autoregressive order of the process with coefficients ϕ_i . Thus, the $AR(p)$ takes into account just the p previous periods, while the “new” part of y_t , not linked to the past, is given by ε_t .

ARX. The ARX model is an extension of AR that includes the time series of covariates $\mathbf{x}'_{k,t}$:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{k=1}^n \beta_k x_{k,t} + \varepsilon_t \tag{3}$$

SARIMA. The seasonal ARIMA $(p, d, q) \times (P, D, Q)_S$, or SARIMA, is a process that takes simultaneously into account two features of the observed time series: the correlation between consecutive values modelled by standard ARIMA and the correlation between observations that are far from each other that captures the seasonality. Formally, the ARIMA part of the model is defined as:

$$y_t^d = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + b_t - \sum_{j=1}^q \theta_j b_{t-j}, \tag{4}$$

where p is the autoregressive order of the process with coefficients ϕ_i and q is the order of the moving average process with coefficients θ_i . Notice that in a standard ARIMA process b_t is white noise, whereas here it is not due to the existence of unexplained correlation that we model as follows:

$$w_t = b_t - b_{t-D}, \\ w_t = \sum_{i=1}^P \Phi_i w_{t-i,S} + \varepsilon_t - \sum_{j=1}^Q \Theta_j \varepsilon_{t-j,S}, \tag{5}$$

where D represents the degree of the integration, P and Q are the seasonal orders of the autoregressive and moving average processes with coefficients Φ_i and Θ_i , respectively, S is the seasonality, and $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Using the lag operator B such that $By_t = y_{t-1}$, then (4) and (5) define the SARIMA $(p, d, q) \times (P, D, Q)_S$ process written in compact form:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \phi_0 + \theta(B)\Theta(B^S)\varepsilon_t. \tag{6}$$

SARIMAX. The SARIMAX model is an extension of SARIMA that includes the time series of covariates $\mathbf{x}'_{k,t}$:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \beta_n \mathbf{x}'_{n,t} + \theta(B)\Theta(B^S)\varepsilon_t. \tag{7}$$

Linear Regression. We specify the classical LR model as follows:

$$y_t^1 = \beta_0 + \beta_1 \mathbf{x}_{n,t} + \varepsilon_t, \tag{8}$$

where the dependent variable y_t^1 is the first order differentiated time series at time t and the covariates $\mathbf{x}_{n,t}$ are the variables at time t lagged of h periods where h defines the forecasting horizon. Despite the fact that LR does not account for the autoregressive component, which is typical in a time series, our specification is

built in such a way that allows us to include a degree of temporal information.

2.3 Forecasting Strategies

We propose two different forecasting strategies with the aim of studying the accuracy of the GDP predictions when we include all the available information at present time. We also assess the magnitude of the precision for different forecasting horizons.

One-Step-Ahead Forecasting

The one-step-ahead forecasting strategy computes the forecast for one quarter ahead. This implies that the train set, that is the data used for the forecast, is reduced by one observation that corresponds to the forecasting horizon, which is our test set, and covariates have one period lag. We run the prediction of the GDP for each quarter of the period from the first quarter of 2019 to the last of 2020. In each forecast the test set moves back by one quarter and the train becomes one quarter shorter. It is important to highlight that the chosen out-of-sample forecasting horizon includes both 1 year of normal times (2019) and 1 year affected by the Sars-COVID-19 pandemic (2020). The forecasting methodology works as follow:

$$\begin{array}{ll}
 \text{Train Set} & \text{Test Set} \\
 \mathcal{Y}_{1,t_2}, \dots, \mathcal{Y}_{T,t_i} & \mathcal{Y}_{T,t_{i+1}} \\
 \mathcal{Y}_{1,t_2}, \dots, \mathcal{Y}_{T,t_{i-1}} & \mathcal{Y}_{T,t_i} \\
 \vdots & \vdots \\
 \mathcal{Y}_{1,t_2}, \dots, \mathcal{Y}_{T-h,t_i} & \mathcal{Y}_{T-h,t_{i+1}}
 \end{array} \quad (9)$$

Multi-Step-Ahead Forecasting

In the multi-step-ahead forecasting strategy predictions are run over the horizon that increases at each forecast. In this set up, the end point of the test period is set fixed to the last quarter of 2020 and the starting point moves back by one quarter each forecast. Both GDP and covariates enter the models with a lag equal to the forecasting horizon. The forecasting methodology works as follow:

$$\begin{array}{ll}
 \text{Train Set} & \text{Test Set} \\
 \mathcal{Y}_{1,t_3}, \dots, \mathcal{Y}_{T,t_3} & \mathcal{Y}_{T,t_4} \\
 \mathcal{Y}_{1,t_4}, \dots, \mathcal{Y}_{T,t_2} & \mathcal{Y}_{T,t_3}, \mathcal{Y}_{T,t_4} \\
 \vdots & \vdots \\
 \mathcal{Y}_{4,t_2}, \dots, \mathcal{Y}_{T-3,t_4} & \mathcal{Y}_{T-2,t_1}, \dots, \mathcal{Y}_{T,t_4}
 \end{array} \quad (10)$$

The maximum length of the forecasting horizon here considered is 12 quarters from the first quarter of 2018 to the last of 2020.

3 EMPIRICAL ANALYSIS

3.1 Data

We measure the economic activity with the seasonally adjusted real U.S. GDP expressed in quarterly frequency and in log scale. The data span the period from second quarter of 1976 to fourth quarter of 2020, for an overall of 179 observations, and are

available from the database of the Federal Reserve Bank of Saint Louis, Federal Reserve Economic Data, FRED.

Interest rates and proxies. Both short-term and long-term U.S. federal government interest rates are used in our study. Short-term interest rates are obtained from Treasury-Bills with maturities 3 and 6 months; long-term interest rates are from the U.S. government bonds with maturities of 2, 3, 5, 7 and 10 years. Drawing on Diebold et al. (2006) and Ang et al. (2006), we exploit an alternative representation of the yield curve through its latent factors, namely the level, slope, and curvature to capture the economic information contained in it. The level is computed taking the average of short-, medium- and long-term bonds; in our study we use the interest rates at 3 months, 2 and 10 years. The slope is the result of the difference between the shortest- and the longest-term yield, 3 months and 10 years. The curvature is estimated computing the double product of the medium-term yield minus the shortest- and the longest-term yield.

Macroeconomic variables. We extend the analysis introducing key observable macroeconomic variables. Following the existing literature (Ang et al., 2006; Diebold et al., 2006; Koop, 2013; Schorfheide and Song, 2015) we select the Consumer Price Index, Manufacturing Capacity Utilization, and Unemployment Rate to illustrate real economic activity whereas the Federal Funds rates proxies the monetary policy. The Manufacturing Capacity Utilization and the Consumer Price Index are differentiated to make the series stationary.

3.2 Models Fitting

KNN. Performing a grid search we find that optimal value of k is 2 for both forecasting strategies.

AR. We use stepwise procedure in order to choose the optimal autoregressive value of p , minimizing the AIC value.

ARX. The same methodology of AR has been applied to ARX.

SARIMA. With quarterly GDP data the seasonal period of the series is $s = 4$. Therefore, (11) becomes:

$$\Phi_p(B^4)\phi(B)\nabla_4^D\nabla^d y_t = \Theta_Q(B^4)\theta(B)w_t. \quad (11)$$

The orders p, d, q and P, D, Q are chosen performing stepwise search to minimize the AIC selection criterion.

SARIMAX. By (11), (7) becomes:

$$\Phi_p(B^4)\phi(B)\nabla_4^D\nabla^d y_t = \beta_n' \mathbf{x}_{n,t}' + \Theta_Q(B^4)\theta(B)w_t. \quad (12)$$

Linear Regression. We fit a linear regression for each scenario and forecasting strategy. In the one-step-ahead forecasts the covariates have one period lag. In the multi-step-ahead the covariates have a lag equal to the length of the forecasting horizon, which increases at each forecast.

We include a set of covariates $\mathbf{x}'_{n,t}$ in LR, ARX and SARIMAX and study six different scenarios:

- Scenario 1 = {Yield Curve};
- Scenario 2 = {Yield Curve, Macro-variables};
- Scenario 3 = {Macro-variables};
- Scenario 4 = {Proxies};
- Scenario 5 = {Macro-variables, Proxies};
- Scenario 6 (Full) = {Yield Curve, Macro-variables, Proxies};

TABLE 1 | Average MSE for all periods.

| Strategy | SARIMA | KNN | AR |
|------------|----------|----------|----------|
| One-step | 2,87e-03 | 1,73e-03 | 3,47e-03 |
| Multi-step | 3,84e-03 | 3,02e-03 | 4,18e-03 |

where the covariates for the yield curve are Treasury-Bills with maturities 3 and 6 months and 2, 3, 5, 7 and 10 years. Macro variables are Consumer Price Index, Manufacturing Capacity Utilization, Unemployment rate, and the Federal Funds rate. The proxies are the level, slope and curvature.

4 RESULTS

4.1 Model Performances

The KNN model achieves the best forecasting results with respect to SARIMA and AR, specifications that do not include covariates, as reported in **Table 1**.

Other models that provide good forecasts are models that include covariates, namely SARIMAX, LR and ARX. We notice that both SARIMAX and LR tend to overestimate the GDP predictions. We also investigate the average of the predictions obtained with the two models (Mean LR-SARIMAX):

$$\hat{y} = \frac{1}{2} (\hat{y}_{t,LR} + \hat{y}_{t,SARIMAX}). \quad (13)$$

Table 2 reports the average MSE. Among all the models, KNN provides the best forecasts. SARIMAX and ARX are able to better predict the GDP one-step-ahead when interest rates (Scenario 1) and proxies (Scenario 4) are considered as covariates. This finding remains true also when forecasting with the multi-step-ahead strategy.

Overall, the one-step-ahead predictions with Scenarios 1 and 4 are the most accurate, whereas the multi-step-ahead forecast with macro variables (Scenario 3) contributes to improve the predictions the most. The Mean LR-SARIMAX performs equally likely as the SARIMAX.

4.2 Out-of-Sample One-Step-Ahead Forecasting Performance

Table 3 displays the prediction accuracy for the forecasting horizon of Scenario 4 (proxies). **Figure 3** shows the accuracy, in terms of MSE, that fitted models achieve in each forecast horizon in the one-step-ahead strategy. The clear pattern that emerges is the change in the best performing model due to the COVID-19 shock. Specifically, models with the autoregressive component perform better before the second quarter of 2020 while the other models result to better respond to COVID-19. On the one hand, the KNN provides the best out-of-sample prediction for the second quarter of 2020 that corresponds to the beginning of the pandemic outbreak. On the other hand, SARIMAX is more accurate in normal periods as it achieves the lowest forecast

TABLE 2 | Comparison of average MSE in the two strategies considered: one-step- vs. multi-step-ahead.

| | SARIMAX | LR | Mean LR-SARIMAX | ARX |
|------------|----------|----------|-----------------|----------|
| One step | | | | |
| Scenario 1 | 2,44e-03 | 1,93e-03 | 2,15e-03 | 3,57e-03 |
| Scenario 2 | 6,19e-03 | 2,41e-03 | 3,96e-03 | 7,81e-03 |
| Scenario 3 | 5,90e-03 | 2,38e-03 | 3,84e-03 | 5,92e-03 |
| Scenario 4 | 2,86e-03 | 1,89e-03 | 2,31e-03 | 3,45e-03 |
| Scenario 5 | 6,10e-03 | 2,43e-03 | 3,96e-03 | 6,66e-03 |
| Scenario 6 | 6,14e-03 | 2,41e-03 | 3,94e-03 | 7,99e-03 |
| Multi step | | | | |
| Scenario 1 | 3,66e-03 | 2,50e-03 | 3,01e-03 | 4,29e-03 |
| Scenario 2 | 3,99e-03 | 2,48e-03 | 3,13e-03 | 5,24e-03 |
| Scenario 3 | 4,05e-03 | 2,38e-03 | 3,08e-03 | 4,69e-03 |
| Scenario 4 | 3,97e-03 | 2,40e-03 | 3,06e-03 | 5,24e-03 |
| Scenario 5 | 4,57e-03 | 2,37e-03 | 3,26e-03 | 4,84e-03 |
| Scenario 6 | 3,97e-03 | 2,48e-03 | 3,13e-03 | 4,86e-03 |

error for the first quarter of 2019. The same holds for both AR and ARX which are the most accurate in the second quarter of 2019. SARIMA is the best performing model for the fourth quarter of 2019.

The second best forecasting model is the LR. As shown in **Figure 3**, it performs well on the whole forecasting horizon. Looking at single scenarios that include the LR outperforms the other models, confirming the forecasting-power of the yield curve in predicting the GDP.

4.3 Out-of-Sample Multi-Step-Ahead Forecasting Performance

Table 4 shows the results of the second type of forecasting strategy for the Scenario 5 (proxies and macro variables). **Figure 4** shows the MSE of the models for each forecasting horizon. The change occurs also for the multi-step-ahead strategy and the time series models lose the most in terms of performance after the second quarter of 2020. The best overall performance is achieved by the LR with this specification. We highlight that such set of covariates performs better than other combinations, namely Scenario 1, 2, 3, 4, and 6. The average MSE with Scenario 5 is the lowest among models with and without covariates. This result holds true for both periods of stability and crisis. A possible justification lies in the fact that the LR does not include the autoregressive term of the GDP that may affect the prediction performance. Indeed, the macro variables may be more reactive improving the prediction compared to autoregressive models.

5 CONCLUSION

In this article, we provide a comparison of the predictive ability of time series, linear regression, and machine learning models to forecast the U.S. GDP. We discuss the benefit for the policymaker of a regression approach compared to the classification to address medium-term policies. Moreover, we evaluate two different

TABLE 3 | MSE One-step-ahead forecast, Scenario 4.

| Dates | SARIMAX | LR | Mean LR-SARIMAX | SARIMA | KNN | AR | ARX |
|------------|----------|----------|-----------------|----------|----------|----------|----------|
| 2020-10-01 | 2,27e-05 | 3,51e-05 | 2,86e-05 | 1,85e-05 | 2,56e-05 | 1,54e-03 | 1,30e-03 |
| 2020-07-01 | 1,35e-02 | 4,97e-03 | 8,72e-03 | 1,36e-02 | 6,14e-03 | 1,77e-02 | 1,78e-02 |
| 2020-04-01 | 9,03e-03 | 9,79e-03 | 9,41e-03 | 9,05e-03 | 7,33e-03 | 8,25e-03 | 8,24e-03 |
| 2020-01-01 | 2,85e-04 | 3,31e-04 | 3,07e-04 | 3,07e-04 | 2,97e-04 | 2,83e-04 | 2,54e-04 |
| 2019-10-01 | 8,94e-07 | 1,81e-06 | 1,31e-06 | 2,02e-09 | 1,28e-05 | 2,76e-06 | 4,19e-06 |
| 2019-07-01 | 1,03e-06 | 2,84e-06 | 1,82e-06 | 7,14e-07 | 3,07e-07 | 6,50e-06 | 6,70e-06 |
| 2019-04-01 | 1,91e-05 | 3,17e-06 | 9,44e-06 | 1,73e-05 | 4,80e-05 | 3,58e-07 | 1,11e-06 |
| 2019-01-01 | 2,10e-07 | 3,27e-07 | 2,65e-07 | 1,23e-06 | 1,75e-06 | 1,57e-05 | 1,70e-05 |

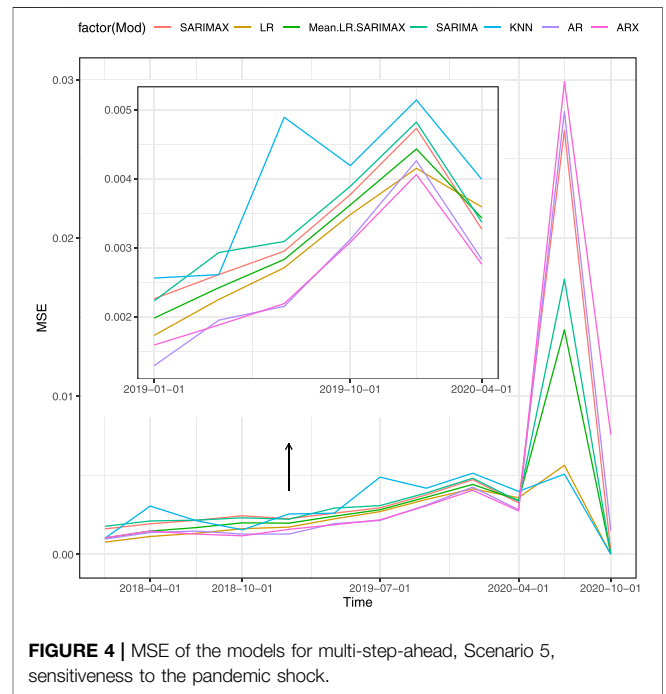
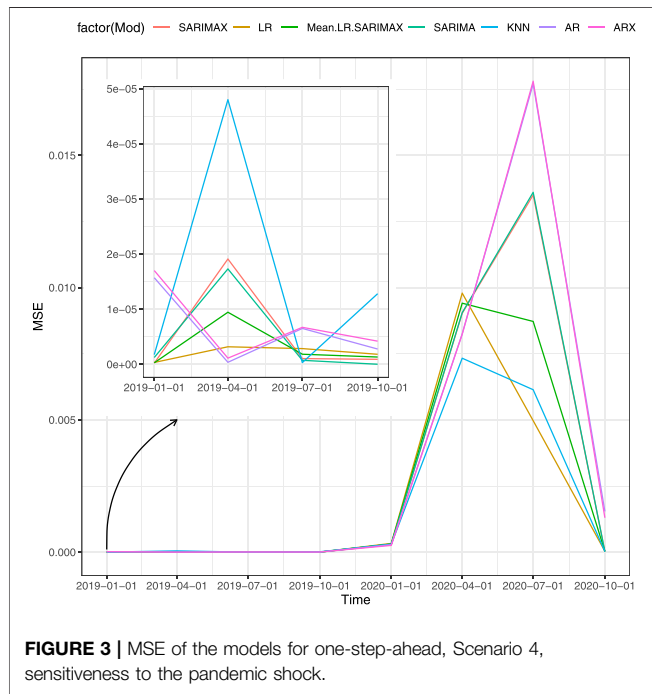


TABLE 4 | MSE multi-step-ahead forecast, Scenario 5.

| Dates | SARIMAX | LR | Mean LR-SARIMAX | SARIMA | KNN | AR | ARX |
|---------------|----------|----------|-----------------|----------|----------|----------|----------|
| 2020q4 | 2,42e-04 | 5,00e-06 | 4,40e-05 | 1,80e-05 | 2,56e-05 | 1,54e-03 | 7,58e-03 |
| 2020q3–2020q4 | 2,68e-02 | 5,64e-03 | 1,42e-02 | 1,74e-02 | 5,08e-03 | 2,80e-02 | 2,99e-02 |
| 2020q2–2020q4 | 3,27e-03 | 3,59e-03 | 3,43e-03 | 3,37e-03 | 3,99e-03 | 2,83e-03 | 2,76e-03 |
| 2020q1–2020q4 | 4,73e-03 | 4,15e-03 | 4,43e-03 | 4,82e-03 | 5,14e-03 | 4,26e-03 | 4,06e-03 |
| 2019q4–2020q4 | 3,77e-03 | 3,48e-03 | 3,62e-03 | 3,89e-03 | 4,19e-03 | 3,12e-03 | 3,08e-03 |
| 2019q3–2020q4 | 2,95e-03 | 2,71e-03 | 2,83e-03 | 3,09e-03 | 4,89e-03 | 2,15e-03 | 2,19e-03 |
| 2019q2–2020q4 | 2,61e-03 | 2,25e-03 | 2,42e-03 | 2,93e-03 | 2,61e-03 | 1,95e-03 | 1,88e-03 |
| 2019q1–2020q4 | 2,26e-03 | 1,73e-03 | 1,98e-03 | 2,23e-03 | 2,56e-03 | 1,29e-03 | 1,59e-03 |
| 2018q4–2020q4 | 2,45e-03 | 1,63e-03 | 2,00e-03 | 2,32e-03 | 1,56e-03 | 1,29e-03 | 1,18e-03 |
| 2018q3–2020q4 | 2,16e-03 | 1,34e-03 | 1,70e-03 | 2,17e-03 | 2,14e-03 | 1,48e-03 | 1,29e-03 |
| 2018q2–2020q4 | 1,94e-03 | 1,14e-03 | 1,48e-03 | 2,12e-03 | 3,06e-03 | 1,39e-03 | 1,48e-03 |
| 2018q1–2020q4 | 1,62e-03 | 7,83e-04 | 1,04e-03 | 1,78e-03 | 1,03e-03 | 9,78e-04 | 1,06e-03 |

strategies of forecasting, one-step-ahead and multi-step-ahead, considering the self-explanatory power of GDP and the importance of financial and macro-economic variables as predictors. On the one hand, the machine learning KNN

achieves the best performance for the one-step-ahead strategy, providing evidence that in the subsequent horizon the exploitation of repetitive patterns in the GDP increases the forecast. On the other hand, it loses predictive power when

the forecast is performed for a longer horizon. SARIMA performs poorly in the one-step-ahead and multi-step-ahead strategies. Including covariates, SARIMAX obtains a lower error in the one-step-ahead strategy especially with the Treasury-Bills with maturities 3 and 6 months and 2, 3, 5, 7 and 10 years (Scenario 1). ARX achieves the best forecasting performance in one-step-ahead with proxies (Scenario 4) and yield curve (Scenario 1). Considering the multi-step-ahead accuracy, the yield curve has proved to be the best predictor to be paired with this model. Surprisingly, the LR achieves the best performance in the multi-step-ahead forecast using proxies for the yield curve and macro variables (Scenario 5). Moreover, it achieves the second-best performance in the one-step-ahead strategy using only the proxies as predictors and confirming the strong predictive power of the yield curve for the GDP. In general, we find that a switch occurs in terms of forecasting performances, both for one and multi-step-ahead (see **Figures 3, 4**), between models which have the autoregressive component and models without it. Before the cutoff, the pandemic outbreak in our study, time series models perform better but after that event LR and KNN outperform the other approaches. The results of our analysis suggest the use of the KNN model for one-step-ahead forecasts and that of LR with the use of financial variables for multi-step-ahead forecasts. We propose to overcome the trade-off between accuracy in the estimates and the forecasting horizon, considering the two forecasting strategies which are not mutually exclusive. Indeed, the benefit of a

continuous forecasting of both one-step-ahead and multi-step-ahead allows the decision-maker to have two useful instruments: on the one hand the multi-step provides a long-term vision for planning in advance investments, monetary policy, etc., on the other hand the one-step-ahead might tip the scale for possible refinement around the decision taken. There are many possible avenues for future works. A desirable address is to develop a model that includes the international bond yield curve (Byrne et al., 2019), macro variables, and the GDP of countries the United States trade with.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are freely available at the Federal Reserve Bank of Saint Louis, Federal Reserve Economic Data (FRED) website, <https://fred.stlouisfed.org/>. The interested reader may use the provided link to FRED to explore the data. Any further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GMa, GMo and SS conceptualized the topic, performed the literature review, drafted, edited and reviewed the article and approved the manuscript for submission. GMa and SS implemented the code.

REFERENCES

- Al-Qahtani, F. H., and Crone, S. F. (2013). "Multivariate K-Nearest Neighbour Regression for Time Series Data A Novel Algorithm for Forecasting uk Electricity Demand," in The 2013 international joint conference on neural networks (IJCNN) (IEEE), 1–8.
- Ang, A., Piazzesi, M., and Wei, M. (2006). What Does the Yield Curve Tell Us about Gdp Growth? *J. Econom.* 131, 359–403. doi:10.1016/j.jeconom.2005.01.032
- Atsalakis, G. S., Bouri, E., and Pasiouras, F. (2020). Natural Disasters and Economic Growth: a Quantile on Quantile Approach. *Ann. Operations Res.* 5, 1–27. doi:10.1007/s10479-020-03535-6
- Baffigi, A., Golinelli, R., and Parigi, G. (2002). *Real-time GDP Forecasting in the Euro Area*, Vol. 456. Princeton, NJ: Citeseer.
- Ban, T., Zhang, R., Pang, S., Sarrafzadeh, A., and Inoue, D. (2013). "Referential Knn Regression for Financial Time Series Forecasting," in International Conference on Neural Information Processing (Springer), 601–608. doi:10.1007/978-3-642-42054-2_75
- Bañbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian Vector Auto Regressions. *J. Appl. Econom.* 25, 71–92.
- Bellotti, A., Brigo, D., Gambetti, P., and Vrins, F. (2021). Forecasting Recovery Rates on Non-Performing Loans With Machine Learning. *Int. J. Forecast.* 37, 428–444. doi:10.1016/j.ijforecast.2020.06.009
- Bernard, H., and Gerlach, S. (1998). Does the Term Structure Predict Recessions? the International Evidence. *Int. J. Fin. Econ.* 3, 195–215. doi:10.1002/(sici)1099-1158(199807)3:3<195:aid-ijfe81>3.0.co;2-m
- Bertsimas, D., Orfanoudaki, A., and Pawlowski, C. (2021). Imputation of Clinical Covariates in Time Series. *Mach Learn.* 110, 185–248. doi:10.1007/s10994-020-05923-2
- Bouheni, F. B., Obeid, H., and Margarint, E. (2021). Nonperforming Loan of European Islamic Banks Over the Economic Cycle. *Ann. Operations Res.*, 1–36. doi:10.1007/s10479-021-04038-8
- Brave, S. A., Butters, R. A., and Justiniano, A. (2019). Forecasting Economic Activity With Mixed Frequency Bvars. *Int. J. Forecast.* 35, 1692–1707. doi:10.1016/j.ijforecast.2019.02.010
- Byrne, J. P., Cao, S., and Korobilis, D. (2019). Decomposing Global Yield Curve Co-Movement. *J. Banking Finance.* 106, 500–513. doi:10.1016/j.jbankfin.2019.07.018
- Chauvet, M., and Potter, S. (2013). Forecasting Output. *Handbook Econ. Forecast.* 2, 141–194. doi:10.1016/b978-0-444-53683-9.00003-7
- Diebold, F. X., and Nason, J. A. (1990). Nonparametric Exchange Rate Prediction? *J. Int. Econ.* 28, 315–332. doi:10.1016/0022-1996(90)90006-8
- Diebold, F. X., Rudebusch, G. D., and Boragan Aruoba, S. (2006). The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach. *J. Econom.* 131, 309–338. doi:10.1016/j.jeconom.2005.01.011
- Doz, C., Giannone, D., and Reichlin, L. (2011). A Two-step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *J. Econom.* 164, 188–205. doi:10.1016/j.jeconom.2011.02.012
- Estrella, A., and Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *J. Finance* 46, 555–576. doi:10.1111/j.1540-6261.1991.tb02674.x
- Estrella, A., and Mishkin, F. S. (1996). "The Yield Curve as a Predictor of Recessions in the united states and Europe," in *The Determination of Long-Term Interest Rates and Exchange Rates and the Role of Expectations* (Basel: Bank for International Settlements).
- Faroni, C., Marcellino, M., and Stevanovic, D. (2020). Forecasting the Covid-19 Recession and Recovery: Lessons From the Financial Crisis. *Int. J. Forecast.* doi:10.1016/j.ijforecast.2020.12.005
- Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The Real-Time informational Content of Macroeconomic Data. *J. Monetary Econ.* 55, 665–676. doi:10.1016/j.jmoneco.2008.05.010
- Hamilton, J. D. (2018). Why You Should Never Use the Hodrick-Prescott Filter. *Rev. Econ. Stat.* 100, 831–843. doi:10.1162/rest_a_00706
- Hodrick, R. J., and Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *J. Money, Credit Banking.* 29, 1–16. doi:10.2307/2953682
- Hu, Z., Ge, Q., Li, S., Boerwinkle, E., Jin, L., and Xiong, M. (2020). Forecasting and Evaluating Multiple Interventions for Covid-19 Worldwide. *Front. Artif. Intell.* 3, 41. doi:10.3389/frai.2020.00041
- Koop, G., McIntyre, S., Mitchell, J., and Poon, A. (2020). *Reconciled Estimates of Monthly GDP in the US. Warwick, United Kingdom: Economic Statistics Centre of Excellence.*

- Koop, G. M. (2013). Forecasting with Medium and Large Bayesian Vars. *J. Appl. Econ.* 28, 177–203. doi:10.1002/jae.1270
- Luginbuhl, R., and Koopman, S. J. (2004). Convergence in European GDP Series: a Multivariate Common Converging Trend-Cycle Decomposition. *J. Appl. Econ.* 19, 611–636. doi:10.1002/jae.785
- Lunde, A., and Torkar, M. (2020). Including News Data in Forecasting Macro Economic Performance of China. *Comput. Manag. Sci.* 17, 585–611. doi:10.1007/s10287-020-00382-5
- Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A Comparison of Direct and iterated Multistep Ar Methods for Forecasting Macroeconomic Time Series. *J. Econom.* 135, 499–526. doi:10.1016/j.jeconom.2005.07.020
- Martínez, F., Frías, M. P., Charte, F., and Rivera, A. J. (2019). Time Series Forecasting With Knn in R: the Tsfknn Package. *R. J.* 11, 229. doi:10.32614/rj-2019-004
- Provost, F., and Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big data.* 1, 51–59. doi:10.1089/big.2013.1508
- Rodríguez-Vargas, A. (2020). Forecasting Costa Rican Inflation With Machine Learning Methods. *Latin Am. J. Cent. Banking.* 1, 100012. doi:10.1016/j.latchb.2020.100012
- Schorfheide, F., and Song, D. (2015). Real-Time Forecasting With a Mixed-Frequency Var. *J. Business Econ. Stat.* 33, 366–380. doi:10.1080/07350015.2014.954707
- Schüler, Y. (2018). *On the Cyclical Properties of Hamilton's Regression Filter*. Discussion Papers 03/2018. Amsterdam: Deutsche Bundesbank.
- Spelta, A., and Pagnottoni, P. (2021). Mobility-based Real-Time Economic Monitoring Amid the Covid-19 Pandemic. *Sci. Rep.* 11, 13069–13115. doi:10.1038/s41598-021-92134-x
- Stone, C. J. (1977). Consistent Nonparametric Regression. *Ann. Stat.* 5, 595–620. doi:10.1214/aos/1176343886
- Taylor, J. B. (1993). Discretion Versus Policy Rules in Practice. *Carnegie-Rochester Conf. Ser. Public Pol.* 39, 195–214. doi:10.1016/0167-2231(93)90009-1
- Wauters, M., and Vanhoucke, M. (2017). A Nearest Neighbour Extension to Project Duration Forecasting With Artificial Intelligence. *Eur. J. Oper. Res.* 259, 1097–1111. doi:10.1016/j.ejor.2016.11.018
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., et al. (2008). Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst.* 14, 1–37. doi:10.1007/s10115-007-0114-2
- Yiu, M. S., and Chow, K. K. (2010). Nowcasting Chinese Gdp: information Content of Economic and Financial Data. *China Econ. J.* 3, 223–240. doi:10.1080/17538963.2010.562028

Conflict of Interest: Author SS was employed by company Enel Global Services S.r.l.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Maccarrone, Morelli and Spadaccini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.