INTERNATIONAL MONETARY FUND

Nowcasting GDP

A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies

By Jean-François Dauphin, Kamil Dybczak, Morgan Maneely, Marzie Taheri Sanjani, Nujin Suphaphiphat, Yifei Wang, and Hanqi Zhang

WP/22/52

2022
MAR

WORKING PAPER

 WP/22/52

IMF Working Paper
European Department

Nowcasting GDP
A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies

Prepared by Jean-François Dauphin, Kamil Dybczak, Morgan Maneely, Marzie Taheri Sanjani, Nujin Suphaphiphat, Yifei Wang, and Hanqi Zhang

March 2022

ABSTRACT: This paper describes recent work to strengthen nowcasting capacity at the IMF's European department. It motivates and compiles datasets of standard and nontraditional variables, such as Google search and air quality. It applies standard dynamic factor models (DFMs) and several machine learning (ML) algorithms to nowcast GDP growth across a heterogenous group of European economies during normal and crisis times. Most of our methods significantly outperform the AR(1) benchmark model. Our DFMs tend to perform better during normal times while many of the ML methods we used performed strongly at identifying turning points. Our approach is easily applicable to other countries, subject to data availability.

JEL Classification Numbers:	E52, C33, C53
Keywords:	Nowcasting, Factor Model, Machine Learning, Large Data Sets
Author's E-Mail Address:	jdauphin@imf.org; kdybczak@imf.org; mmaneely@imf.org; nsuphaphiphat@imf.org; mtaherisanjani@imf.org; ywang6@imf.org; hqzhang@nyu.edu

WORKING PAPERS

Nowcasting GDP

A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies

Prepared by Jean-François Dauphin, Kamil Dybczak, Morgan Maneely, Marzie Taheri Sanjani, Nujin Suphaphiphat, Yifei Wang, and Hanqi Zhang [1]

[1] The author(s) would like to thank participants of the IMF Big Data Talks and European Department seminars for suggestions. All errors and omissions are our own.

Contents

Glossary 3

Introduction 4

Literature Review 5

Data 7

Models and Methodology 9
- DFMs 10
- Machine Learning 11
- Assessment of Predictive Performance 12

Results 14

Summary and Conclusion 18

Appendices 20
- 1. Nowcasting Variables by Country 20
- 2. Sample Results—Hungary 27
- 3. An Integrated Tool 32
- 4. Machine Learning Algorithms 34
  - 1. Regularized Regression methods 34
  - 2. Support Vector Machine 34
  - 3. Random Forest 36
  - 4. Neural Network 38
- 5. Indicators of predictive accuracy: Models by Country 40

References 41

FIGURES
- Figure 1. Examples of Model Performances 16

TABLES
- Table 1. A Brief Introduction to ML Algorithms 12
- Table 2. Full Sample- Model Performance (RMSE) Until 2021Q1 17
- Table 3. Pre-COVID Samples- Model Performance (RMSE) Until 2019Q4 17
- Table 4. During COVID-19 Sample- Model Performance (RMSE) Between 2020Q1-2021Q1 18

Glossary

AR	Auto Regressive
CNN	Convolutional Neural Network
COVID	Coronavirus Disease
CPI	Consumer Price Index
DFM	Dynamic Factor Model
ECB	European Central Bank
EM	Expectation Maximization
GDP	Gross Domestic Product
GPReg	Gaussian Process Regression
IMF	International Monetary Fund
LASSO	Least Absolute Shrinkage and Selection Operator
LinReg	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MDA	Mean Directional Accuracy
MIDAS	Mixed Data Sampling
ML	Machine Learning
NN	Neural Network
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
ReLU	Rectifier Linear Unit
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recursive Neural Networks
SVM	Support Vector Machine
VAR	Vector Autoregressive Model
WEI	Weekly Economic Index

Introduction

The Covid crisis highlighted the need for a systematic analysis of high-frequency indicators of economic activity. While in normal times the dynamics of real economic activity can generally be sufficiently characterized and explained by conventional monthly and quarterly aggregate indicators—e.g., industrial production or gross domestic product— during crisis times, businesses, market analysts, and policymakers need higher-frequency information about economic activity. The importance of timely information was vividly illustrated in 2020, when the COVID pandemic hit the global economy with unprecedented intensity and unpredictability. In Europe for instance, the economy recorded in the second quarter its largest contraction since World War II. However, the first official GDP growth estimates of the second quarter for European countries were published only in mid-July, too late to assess the magnitude of the shock and calibrate the appropriate policy response. Fortunately, other economic indicators had been available much earlier. A question arose as to whether a variety of higher frequency data from different sources could be systematically analyzed to provide sufficiently accurate information on economic conditions timely enough to usefully inform policy making.

Over the last decade, at least in advanced countries, data availability has become less of a constraint thanks to new technologies, more efficient at collecting and compiling data. As a result, new information, mainly from non-official sources—such as market surveys, sentiment indicators, Google search results, air quality, etc.—have become available. These indicators—frequently available and possibly capable of quickly registering sharp changes in economic conditions—can supplement standard economic indicators produced and validated by national statistical offices, to help gauge the economic situation at the aggregate level.

Furthermore, methods that can deal with large datasets containing highly correlated variables of different frequencies and information quality, have also been developed. For more than 20 years, models based on dynamic factor methodology have been prominent nowcasting tools to quantitatively assess the state of the economy in the current period or over the very near future. More recently, and not surprisingly, machine learning methods have gained popularity among economists and have been deployed also in nowcasting.

This paper describes an effort to strengthen nowcasting capacity at the IMF's European department in early 2020. It adds to the growing literature on nowcasting in several ways. First, it motivates and compiles datasets of standard variables and new ones, such as Google search and air quality, including country-specific databases of selected European economies and their transformations. Second, beyond applying standard dynamic factor model (DFM), we employ several machine learning (ML) algorithms to nowcast GDP growth. Third, the paper compares the performance of alternative methods across an heterogenous group of European economies, during normal and crisis times.

The paper is structured as follows. Part 2 provides a non-exhaustive literature overview focusing mostly on DFMs and machine learning methods. Part 3 discusses data that could inform about an economy's position and their statistical treatment. Part 4 provides a non-technical summary of DFMs and selected machine learning approaches. Part 5 presents our main results and compares them across alternative techniques, including during normal and crisis episodes. Part 6 concludes.

Literature Review

Economists in academia and policy institutions have sought to predicting quarterly GDP and turning points of the economic cycle over the near term using higher frequency indicators for a long time. Nowcasting models, as we know them today, rely on the idea originally formulated by Sargent and Sims (1977) that the salient features of business cycle fluctuations can be captured by a small number of factors that could be estimated from data. Stock and Watson (1989) were among the first to construct indices of business cycle using factor models.

The early models had the state-space representation and were estimated by maximum-likelihood and relied on only a limited number of variables. The latent common factors were estimated by Kalman Filter. The first model by Stock and Watson (1989) extracted a single common factor from a small set of monthly core indicators (employment, industrial production, sales, and income).

Over time, these models were expanded and improved to deal with a much larger number of variables and factors used to construct different business cycle indicators. Giannone et al. (2005) and Watson (2005) showed that the findings by Sargent and Sims (1997) also apply to high-dimensional datasets and proposed to formulate factor models using state-space specification to reduce the dimensionality of large economic systems, i.e. to extract unknown common factors. While the large number of variables doesn't impose restriction on feasibility of estimation (Doz et al (2012)), it increases the estimation time; hence the need for selectivity when it comes to picking what variables go into a model.

Since then, factor models have been further refined and become widely used tools to monitor economic conditions. The model of Giannone, Reichlin, and Small (2008) was first implemented to nowcast GDP at the Board of Governors of the Federal Reserve based on a project that started in 2003. Various versions of this model have been built for different economies and implemented in other central banks, including the European Central Bank (ECB, 2008) and the International Monetary Fund (Matheson (2011), Taheri Sanjani (2013)). The models have been continuously refined. In a series of papers, D'Agostino et al. (2011) and Giannone et al. (2013) further developed the DFM methodology and improved prediction properties of this group of models by further improving the Kalman-Filter estimation methodlogy and optimizing the data selection process.

DFMs have been widely used during the COVID-19 pandemic. Sampi and Jooste (2020) introduced Google mobility indicators into the DFM model to capture the economic effects of the pandemic on industrial production in Latin America and the Carribean while Lewis et al. (2020a, 2020b) produced a Weekly Economic Index (WEI) to measure real activity. Other work includes Guérin (2020), Chapman and Desai (2021), Antolin-Diaz et al. (2021), or Corona et al. (2021).

While DFMs have became mainstream tools to nowcast GDP growth, new techniques have emerged, based on statistical learning and machine learning. These techniques not only introduce alternative ways to estimate coefficients but also perform data driven variables and model selection that may lead to smaller deviations between actual and predicted outcomes, i.e. better out-of-sample prediction. About thirty years ago, Fagan and Fell (1991) did not include non-linear methods and neural networks in their overview of techniques for short-term economic forecasting, since they were still in their infancy as far as economics was concerned and not widely used by practitioners. Since then, a lot has changed and, while this is a relatively new area, the number of articles related to machine learning using big data in economics has been growing fast due to two factors.

- *First, new data have became available*. For example, Google trends data have become a new source of information. Choi and Varian (2009a) showed that using specific keywords, Google trends data may help predict the present, especially when economic conditions change rapidly. For example, Varian and Choi (2009b), Tuhkuri (2016) used Google search data to predict current unemployment.
- *Second, a large number of algorithms have been developed and improved*. Hastie et al. (2009, and 2015) or Gareth et al. (2015) are classic textbooks discussing the main tools. However, the examples they provide are mostly outside of economics.

As a result, machine learning techniques have become popular among economists. Varian (2014), Belloni et al. (2014), and Storm (2020) provide an overview of some of the tools for manipulating and analyzing big data in economics. Amat et. al. (2018) apply machine learning methods to predict floating exchange rates of twelve industrial countries and conclude that these methods perform better than tools typically used in the literature. Similarly, using a large scanner panel dataset with highly correlated variables, Bajari et al (2015) apply alternative machine learning methods to estimate consumer demand and find that the machine learning methods predict out-of-sample demand much more accurately than standard panel data and logistic models. Chakraborty and Joseph (2017) discuss machine learning in the context of central banking and policy analyses. Beyond discussing main machine learning concepts and models, they also present three case studies using machine learning at the central bank. Kohlscheen (2021) examines drivers of inflation relying on regression trees and random forests. Basuchoudhary et. al. (2014) motivate and apply different machine learning algorithms (Artificial Neural Network, Boosting, Bootstrap Aggregating, Random Forest predictors and Regression Tree) to predict long-term economic growth and the likelihood of recessions. They use the machine learning algorithms to assess the significance of various variables —previously identified by research—in predicting economic growth. Athey (2019) compares machine learning methods to traditional econometrics, discusses main advantages of machine learning, and provides examples of using these methods for prediction and inference in policy analysis. She also predicts that machine learning will have a large impact on how empirical work is conducted.

Although machine learning methods have been primarily developed for prediction, they have started being applied also for nowcasting. Richardson et. al. (2020) apply machine learning algoriths to a dataset of about 600 domestic and international variables to predict New Zealand GDP growth. The study shows that machine learning algorithms—mainly boosted trees, support vector machine, and neural networks—are able to outperform a simple autoregressive model and DFM. Applying several machine learning algorithms on a dataset of quarterly macroeconomic and financial data to nowcast Indonesia's GDP growth, Muchisha et. al. (2021) arrive at a similar conclusion that all machine learning models outperform AR(1) benchmark while Random Forest showed the best performance. Without adding data from more granular and novel data sources, Jung et. al. (2018) still find that machine learning methods—Elastic Net, SuperLearner, and Recurring Neural Network algorithms—can outperform traditional statistical methods. Tiffin (2016) applies machine learning algorithms in case of Lebanon—a country with long delays in the publication of its GDP. Using a dataset of 19 variables, he shows how Elastic Net and Random Forest can be used to nowcast Lebanese GDP growth before official data is released. Bolhuis and Rayner (2020) apply alternative machine learning algorithms to nowcast the Turkish GDP growth. In addition, to further lower forecast errors, the authors combine individual machine learning models into ensembles. They find that the machine learning models can reduce forecast errors by at least 30 percent when compared to traditional models. In response to rapidly changing environmend during the COVID pandemic, Woloszko (2020) constructed the OECD Weekly Tracker of GDP growth by including Google trends data highlighting the predictive power of specific keywords, including "bankruptcies", "economic crisis", "investment", "luggage" and "mortgage". On a sample of OECD and G20 countries, they find that on average the model's forecast error is 17 percent lower than that of an AR(1) model.

The model also captures a sizeable share of business cycle variation and detects well the downturn and subsequent rebound of the economy.

In addition to DFM and machine learning, nowcasting models can be based on other approaches, including for example bridge equations, mixed data sampling (MIDAS) or mixed-frequency VARs. *Bridge equations* gained in popularity mainly due to its simplicity and low technical requirements. The method relies on single equation regressions of quarterly GDP growth on a small number of high-frequency variables (such as industrial production or surveys) aggregated to quarterly data by averaging, summing up, or taking another transformation (Baffigi et al., 2004). To form a forecast of the variable of interest, high-frequency (explanatory) variables are forecast using separate time series models with a potentially large numbers lags, resulting in a large number of parameters to be estimated. Thus, while parsimonious, the main disadvantage of bridge equations models is a potentially large number of parameters that needs to be estimated. *MIDAS* approach addresses this issue, e.g., the large number of parameters that needs to be estimated, by replacing the unrestricted lag polynomial structure by non-linear functional forms. While MIDAS approach achieves the model's parsimony, its parameters need to be estimated by non-linear estimation methods, which may complicate empirical work. *Mixed frequency VARs* are an alternative to single-equation models. These models operate at the highest frequency of data, at which, however, not all variables are necessarily available. Lower frequency variables are assumed to be periodically missing and represented through a state space model.

Over the last decade, these and other new approaches have been introduced to near-term forecasting and nowcasting literature, all of them having their pros and cons (Camacho and Perez-Quiros (2013)). The remainder of this paper will focus on nowcasting models based on DFMs and machine learning algorithms.

Data

The basic idea of nowcasting is to exploit a diverse set of timely information that is available before an official release of a target variable.[1] As such, data selection and transformation is key to the success of nowcasting.

New technologies and approaches to data collection contributed to wider data availability, allowing economists to rely on large datasets. The so-called high-dimensional datasets or big data often include a number of explanatory variables that is close to, or even exceeds, the number of observations. The motivation behind those large datasets has been to maximize the information set and thus reduce the risk of bias due to the omission of information.

Unfortunately, the prediction capacity of a fitted model does not always improve when introducing additional explanatory variables (Hastie et al (2015)). Specifically, introducing additional explanatory variables into a model increases the risk of overfitting the model—as potentially noisy variables are assigned non-zero coefficients. In addition, even if explanatory variables are relevant to the model, large datasets often consist of interconnected clusters of information, within which variables are highly correlated or present similar

[1] Nowcasting methods have been successfully applied in a number of fields, including outside of economics. In economics, nowcasting methods have been applied not only in near-term economic activity prediction but also to nowcast inflation or employment (Knotek and Saeed, 2014; Modugno, 2011; Hutter, 2020). The focus of this work is nowcasting of quarterly GDP growth.

concepts.[2] In general, excessive collinearity among explanatory variables results in the large variance of the estimated coefficients, which may overweight the benefits of reduced omitted variable bias.

To address issues arising from large datasets—which have become limiting especially in case of classical statistical techniques such as linear regression—and to fully exploit their potential, new techniques and algorithms have been developed. Recently developed techniques and algorithms have been instrumental in dealing with large datasets (see Models and Methodology section). Some of the models penalize the introduction of additional variables and some of them have the capacity to fully exclude variables from the sample and thus shrink the size of the original dataset. While this is an important feature that makes the practitioner's task easier, carefully selecting timely and informative variables has remained critical to nowcasting in the big data environment.

Following standard practices in nowcasting of GDP growth, our sample contains a wide range of official and nonofficial, hard and soft, and proxy data. The data used are published with different frequencies and cover main aspects of the economy and the external environment.[3] The dataset does not include detailed sectoral data and uses only headline figures. For example, industrial production is included only once on the aggregate level, although disaggregated data are available at the same time. This practice is in line with the literature showing that disaggregated information neither improves, nor harms the quality of the nowcast.[4]

The dataset also excludes variables available only since 2020. Proxy of economic activity available at frequency close to real time (e.g., Google mobility indicators, dining reservations, hotel occupancy rates, etc.) have gained in popularity over the last decade, and the COVID pandemic has accelerated their use in short-term forecasting. However, while available with high frequency, the time series of most of these indicators are often very short and thus does not allow to estimate, train, and test the performance of alternative models on historical data.[5] As such, Google trends, Oxford Stringency Indices and air quality data are the only high (weekly) frequency variables included in our sample.

The total number of variables ranges between 20 to 60 depending on the characteristics of the economy for which our set of models is applied, and data availability. Data can be organized in the following clusters:

i. *Firm and production data*: industrial output, housing and construction indicators such as manufacturing production and industry turnovers.

ii. *Surveys or forward-looking indicators:* business surveys, economic sentiment survey, and consumer confidence.

iii. *Labor market variables:* unemployment, wages, and employment data.

iv. *Financial variable:* stock prices, various interest rates, stock of credit and deposits

[2] For example, a nowcasting database can consist of many clusters comprising labor market variables, surveys and sentiment indicators, external sector variables (trade and openness), financial and monetary variables, and fiscal variables among others.

[3] Appendix 1 provides a list of variables used for each country included in our exercise. Each list contains individual variables description, publication frequency and the adopted transformation.

[4] Experimenting with a dataset of around 200 time series, Giannone et al. (2008) and Banbura and Modugno (2010) and Banbura et al. (2011) conclude that the marginal impact of disaggregated data on the nowcast precision is minimal. At the same time, these studies also show that including disaggregated data does not deteriorate the performance of the model.

[5] For example, Google Mobility Indicators are available only since the beginning of the Covid pandemic. Moreover, the publication of these indices may be discontinued once the pandemic is contained, and mobility settles in (potentially new) normal levels.

v. *Prices:* producer price index, CPI, house prices, commodity price index

vi. *Foreign trade:* exports and imports volumes and values.

vii. *External environment:* openness to trade, reliance on Global Value Chain, and main trading partners' economic variables.

viii. *Other variables capturing specific features of an economy*: reliance on tourism [6], commodity export, vulnerabilities to climate change, presence of domiciled multinational companies in a country among others.

ix. *Additional indicators*: air quality index, hotel occupancy rates, Oxford Stringency Indices, frequency of google searches for specific words that may proxy current position in the business cycle.[7, 8]

Before applying the nowcasting models, the variables need to be transformed:

- *First, all variables with frequency higher than quarterly (mostly weekly and monthly) must be transformed into quarterly data*. The DFM takes advantage of the mixed-frequency data by using bridge equation and Kalman filter, while in the case of the ML models, weekly and monthly data are transformed to quarterly frequency, so all the variables in the final dataset are on quarterly basis.[9]
- *Second, as in the case of standard econometric tools, time series must be seasonally adjusted and detrended.* In general, transforming quarterly variables in levels into year-over-year growth rates addresses both seasonality and non-stationarity of the time series in the dataset. A smaller number of variables, such as unemployment rate—already expressed as percent—or outcomes from surveys are either kept unchanged in levels or transformed to year-over-year differences.[10]

Models and Methodology

Once the data have been collected and transformed, we need methods to (i) extract information contained in a large number of variables and (ii) address issues arising from large datasets, e.g. collinearity and risk of overfitting. The methods also need to be able to deal with incomplete data and data of different frequencies. The goal is to be able to process new information in real time in a systematic and objective manner.

The remainder of this section discusses main ideas behind techniques applied in this work: (i) DFMs, and (ii) the tools relying on machine learning.[11]

[6] In countries where tourism represents a large share to GDP, like Malta, tourism-related variables such as tourist arrivals and numbers of overnight stays have been incorporated in the dataset.

[7] Air quality data for European countries can be downloaded from the European air Quality Portal. https://aqportal.discomap.eea.europa.eu/

[8] For example, the word unemployment benefit "Arbeitslosengeld" is used to nowcast GDP growth for Austria. In case of Hungary and Portugal "álláskeresési járadék + munkanélküli segély" and "Subsídio de desemprego" have been used, respectively.

[9] Alternatively, weekly and monthly time series can be converted into twelve and three quarterly time series, respectively, the so-called "blocking" (Bańbura et. al., 2013). In our experience with alternative models over different periods, aggregating weekly and monthly data into a single quarterly time series does not lead to lower quality of the forecast due to a loss of information but results in a shorter computational time due to a smaller number of time series.

[10] Appendix 1 presents country specific databases and indicates the transformation applied to each variable.

[11] Appendix 2 and 3 provide additional discussion of DFM and some of the ML based models.

DFMs

DFMs are based on a fundamental idea that a large number of economic variables (N) show similar developments over the business cycle and thus can be described by a small number of common factors (r), i.e., r<N.[12]

Formally, a DFM assumes that many observed variables ($y_{1,t}, \ldots, y_{n,t}$) are driven by a few unobserved dynamic factors ($f_{1,t}, \ldots, f_{r,t}$), while the features that are specific to individual series, such as measurement errors, are captured by idiosyncratic errors ($e_{1,t}, \ldots, e_{n,t}$):

$$y_{i,t} = \lambda_{i,1} f_{1,t} + \cdots + \lambda_{i,r} f_{r,t} + e_{i,t}, \qquad \text{for } i = 1, \ldots, n \quad (1)$$

which relates the data $y_{i,t}$ to the r latent common factors $f_{1,t}, \ldots, f_{r,t}$ through the factor loadings $\lambda_{i,1}, \ldots, \lambda_{i,r}$. The common component $\sum_{j=1}^{r} \lambda_{i,j} f_{j,t}$ represents the part of variables ($y_{1,t}, \ldots, y_{n,t}$) explained by the common factors, while the idiosyncratic component $e_{i,t}$ captures the movements specific to each variable i.

The variable reduction from N to r in the context of DFM models can be done in two ways: (i) naïve or (ii) structured. In the *naïve case*, a factor model agnostically extracts top factors (largest eigenvectors) in the information space so that they jointly explain about 80 percent of variation in the data. In the *structured case*—using prior expertise—the dataset is split into a few clusters and the factor model extracts one factor (the most relevant eigenvector) per cluster.[13] Once factors are collected, the dimension of the data is reduced to only a handful of indices, the so-called factors.

To conduct inference in DFMs using likelihood-based methods and Kalman filtering techniques, the common factors and the idiosyncratic components are modeled as Gaussian autoregressive processes, which account for their serial correlation and persistence.

$$f_{j,t} = a_j f_{j,t-1} + u_{j_t}, \qquad U_{j,t} \sim N\left(0, \sigma^2_{u,j}\right) \quad (2)$$

$$e_{j,t} = \rho_i e_{i,t-1} + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma^2_{\epsilon,t}) \quad (3)$$

Equations 1, 2, and 3 form a state space model where the common factors and the idiosyncratic components are unobserved states. Equation 1 is known as measurement equation and links the data to the unobserved states. Equations 2 and 3, known as the transition equations, describe the dynamics of the system.

The parameters of the model are estimated iteratively using the Kalman smoother and the expectation maximization (EM) algorithm[14]. In the first step, the EM algorithm computes principal components. Then the model parameters are estimated by OLS regression, treating the principal components as if they were the true common factors. This is a good initialization given that principal components are reliable estimates of the common factors when relying on big data. In the second step, using the parameters estimated in the first step, an updated estimate of the common factors is obtained using the Kalman smoother, following Giannone et al. (2008) and Doz et al. (2011). Considering the uncertainty that the factors have been estimated in each round, the maximum likelihood estimate is obtained by iterating the two steps until convergence.

[12] The dynamic factor model used and described in this section closely follows the NY Fed block structure modeling and estimating framework as described by Bok et. el (2017) and Chapter 4 of handbook of economic forecasting volume 2A, Banbura et el (2013).

[13] Section Data splits datasets into clusters.

[14] The expectation maximization algorithm follows the approach described by (Banbura and Modungo, 2010).

Machine Learning

ML algorithms have become widely used in analyzing large datasets for several reasons:[1]

- *Flexibility:* ML algorithms are flexible and effective in capturing patterns in the data. As such, ML techniques often outperform traditional techniques at learning from large datasets and predicting complex and nonlinear relationships.
- *Variable selection:* Most of the ML algorithms limit overfitting by actively and directly seeking the most suitable set of indicators to predict target variables.
- *Cross-validation*: ML algorithms fine-tune the model parameters by iteratively partitioning the sample into training and testing subsamples and gauging the pseudo out-of-sample performance with test subsample to minimize the forecast error.

On the other hand, the research on the interpretability of ML algorithms has been scarce so far (Carvalho et al 2019) and ML algorithms are often seen as black boxes with a clear trade-off between model performance and interpretability (Linardatos et al 2020). In addition, some ML algorithms seem less robust to missing observations or incomplete dataset than DFM.[2]

[1] ML methods have been initially developed for forecasting. Nonetheless, significant progress has been made in applying ML techniques for inference. The focus of this study is ML techniques for prediction only.

[2] A detailed discussion on handling missing data can be found at Marlin (2008). In our toolkit, we use a naïve approach, a combination of either deleting rows or columns with missing data.

Table 1. A Brief Introduction to ML Algorithms	
Methods	**Description**
LASSO, Ridge, Elastic Net	LASSO (least absolute shrinkage and selection operator), Ridge regression and Elastic net are modified linear regression methods introducing different types of regularizations (a penalty imposed on the use of coefficients) to enhance the prediction accuracy. Compared to traditional regression methods, these methods can avoid dimensionality and overfitting, but still face the challenge of linearity.
Support Vector Machine (SVM)	SVM constructs hyperplanes to partition combinations of explanatory variables and makes a point forecast for each of the sections, like kernel regression with regularization. SVM can overcome the drawbacks of linear regression models including linearity, collinearity, overfitting, and high dimension issue. The performance, however, depends on proper selection of the kernel function and regularization parameters. Complicated kernel function or parameters on the other way may limit SVM's interpretability.
Random Forest (RF)	RF is a combination of forecasts from many individual regression trees. As a non-parametric algorithm, RF can also overcome the main drawbacks of linear regression models, including linearity, collinearity, overfitting, and issues related to high dimensionality of datasets. Nonetheless, while flexible, RF has a limited capability of predicting extreme or outlier events. Besides, given the complex structure of RF, it lacks interpretability.
Neural Network (NN)	A typical NN is a multi-layer non-linear method which maps a series of inputs into a target output. A network consists of layers composed of artificial neurons (or nodes). Each neuron (i) receives inputs from neurons in previous layers, (ii) applies a function to produce a single output, and (iii) sends the output to other neurons in the next layer. Eventually, a final set of nodes is mapped into the target output. Being very flexible, NN allow for alternative functional forms in each neuron as well as for different structure of individual layers. As a sophisticated and flexible algorithm, NN has been proven to be an immensely powerful tool for prediction, addressing drawbacks of the traditional regression methods. Such sophistication and flexibility, however, significantly limit the interpretability of predictions from NN.

Assessment of Predictive Performance

Different techniques can perform better than others under specific circumstances. On the one hand, some models may better capture large changes and thus better predict turrning points such as crises. On the other hand, other models may better filter out noise and thus better perform in a more stable environment. Not surprisingly, the DFM and the various ML models may produce results that send contradictory and unclear signals even if applied to an identical dataset.

To determine the predictive accuracy of individual models on ex-post data, we backtest the models and quantify indicators of predictive accuracy using the residuals from the backtest.

- *Backtesting.* In the first stage—to proxy a real-time environment—in each period *t*, data on all explanatory variables are available from the beginning of the sample till the latest quarter *t*. However, information on the target variable is available only up to the *t-1* quarter. In each quarter *t*, each model's parameters are estimated using data up to quarter *t-1*. Afterwards, the estimated models are applied to explanatory variables in time *t* to produce a one-step-ahead out-of-sample prediction (*pseudo* nowcast) of the target variable. Finally, forecast errors are quantified by comparing actual and estimated values of the target variable.
- *Quantification of indicators of predictive accuracy.* In the second stage, using the forecast errors from backtesting, a number of different statistics can be quantified to assess how closely the forecasted variable track the actual data. Since, each indicator have certain properties, e.g., some may put more weight on outliers while others on main trends, we use the following statistics:

  Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{N}}$$

  Mean absolute error (MAE):

$$MAE = \frac{\sum_{t=1}^{N} ABS(y_t - \hat{y}_t)}{N}$$

  Mean Directional Accuracy (MDA):

$$MDA = \frac{1}{N}\sum 1[sgn(y_t - y_{t-1}) = sgn(\hat{y}_t - y_{t-1})],$$

where $1_A(\mathrm{x}) = \begin{cases} 1 \; if \; x \in A \\ 0 \; if \; else \end{cases}$ and $sgn(\mathrm{x}) = \begin{cases} -1 \; if \; x < 0, \\ 0 \; if \; x = 0, \\ 1 \; if \; x > 0 \end{cases}$

  The first two indicators (RMSE and MAE) represent measures of the average model prediction error. By squaring the error terms, RMSE puts higher weights to large errors. It is a useful indicator to compare predictive power of different models particularly when large errors are present. While MAE is only slightly different in definition from the MSE, it has different properties as it assigns same weights to both large and small errors. Thus, unlike the MSE, MAE is not putting too much weight on outliers and it provides a generic measure of how well a model is performing. The MDA is a measure of predictive accuracy that indicates the probability with which a model forecast the actual direction of a time series. In other words, MDA compares the forecast direction (upward or downward) to the actual realized direction.

These statistics of predictive accuracy not only inform practitioners of how closely individual models predict actual data but can also be used to derive weights when aggregating results from individual models. For example, we have used the inverse RMSEs of each model as their weights to produce a weighted average of their predictions.

A specific method may outperform others on a a particular data set, however, other methods may work better on a different data set.[3] Similarly—in the context of nowcasting—a method may perform better during specific episodes, i.e., stable growth or a sharp contraction/revival in economic activity. As there is no criteria for selecting a model a priori based on an individual country's characteristics, selecting a proper method—for any given set of data—that produces the best results is a critical step.To assess and compare the performance

[3] Hastie (2009) or Gareth (2015) provide a more detailed discussion of this issue.

of alternative nowcasting tools during specific episodes, we perform the backtest and quantify the indicators of predictive accuracy for all tools over: (i) *the full sample* as well as (ii) during *normal times* (2015 to 2019Q4), and (iii) during *the period of heightened volatility* triggered by the COVID-19 pandemic (2020Q1 up to 2021q4).

Results

We applied the DFM and ML models to a set of six European countries and examined whether some of the models tend to outperform our benchmark model—AR(1)— and the other models.[4]. Tables 2 to 4 present RMSE of individual nowcasting models and the AR(1) benchmark model.[5] The tables provide the results for the six above mentioned countries during three sample periods. Table 2 covers the *full sample* of available data up to 2021Q1 while Table 3 and 4 provide results for the pre-COVID *normal* period (data up to 2019Q4) and the *COVID pandemic* period (2020Q1 to 2021Q1), respectively. Chart 1 illustrates the variety of fit quality we obtained across countries, models, and time periods. The findings from this exercise are specific to our sample of countries and should not be generalized.

The key findings from the results include the following:

- ***Both our DFM and ML models add value, especially during times of hightened volatility***. Except for Ireland and Malta, the DFM and ML models outperform AR(1) in a large majority of cases when applyied to the whole sample (Table 2). The AR(1) performs quite well during a period of relatively stable growth and frequently outperforms the other nowcasting models (Table 3). For example, the AR(1) benchmark model performed better than several ML methods during normal times in case of Austria, Hungary, Malta, and Poland. However, and not surprisingly, the strength of the DFM and ML models improved markedly during the COVID-19 pandemic (Table 4). In particular, except for Ireland and Malta, the Ridge regression and Lasso models could significantly reduce the average forecast errors ranging from 40 to 75 percent relative to the AR(1) benchmark during the pandemic. For DFM, the forecast errors are about 50 percent lower than the AR(1) benchmark across all countries.

- ***The performance of individual models greatly differs across sub-samples (time periods).*** For example, Ridge Regression and DFM perform well for Poland and Hungary, respectively, during normal times but did not do well during COVID-19 period. Chart 1—middle panel "Not So Good"—provides selected country examples of methods with differing performance during pre- and during COVID periods. Tables 3 and 4 show a full set of the RMSEs for pre- and during COVID periods. As expected, the RMSEs during normal times (pre-COVID) are lower than that during the COVID period, where all countries experienced highly volatile economic growth. Focusing on the period prior to the pandemic, Table 3 shows that the DFM performs relatively well—and often better than many ML model—in case of many countries and is among the best performers for Austria, Hungary, and Portugal. Nonetheless, the DFM model performed worse than the AR(1) benchmark for Malta and Poland. For ML methods, Gaussian Process Regression (GPReg) generates the lowest RSMEs for Ireland and Portugal during normal times, while Linear Regression (LinReg) performs the best for Poland. The results are quite different when considering nowcasting performances during the COVID period (Table 4). Particularly, several ML methods appeared to perform better at capturing

[4] At the time of the writing of this paper, the group of countries have increased from six to 9, including Bulgaria, Russia, and Slovenia.

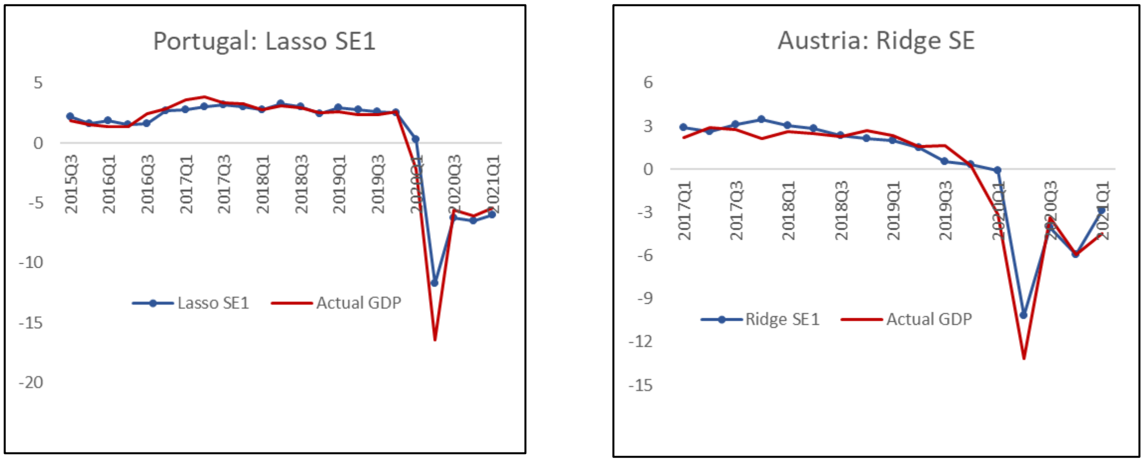
[5] The conclusions presented in this section are based on RMSE solely. Appendix 4 provides additional results for MAE and MDA.

turning points—notably during 2020Q2-2020Q3—when comparing to both the AR(1) benchmark and the DFM. This finding is in line with several papers, including Richardson, Mulder, and Vehbi (2018); and Soybilgen and Yazgan (2021).

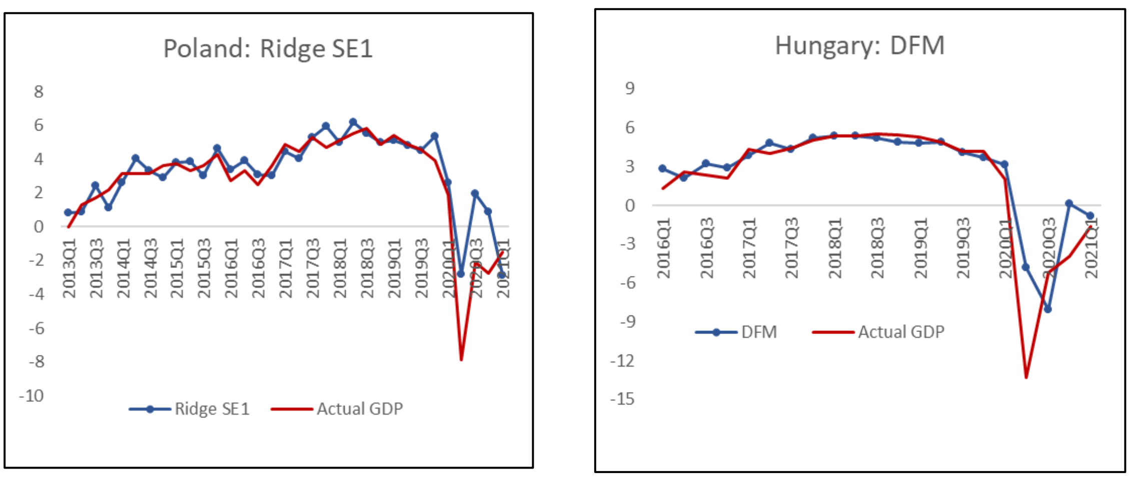
- ***The performances of individual DFM and ML models can also greatly differ across countries***. Tables 2-4 and Panel Chart 1 demonstrate that there is no one-size-fit-all model. In other words, the best model that captures GDP dynamics are different across countries and across time. For example, on the one hand, the DFM generates the lowest RMSE for Malta, while Ridge and LASSO perform the best for Austria, Hungary, and Poland. On the other hand, LASSO does not track well actual GDP growth in case of Malta and Ireland.

Figure 1. Examples of Model Performances

The Good



The (Not So) Bad



The Ugly

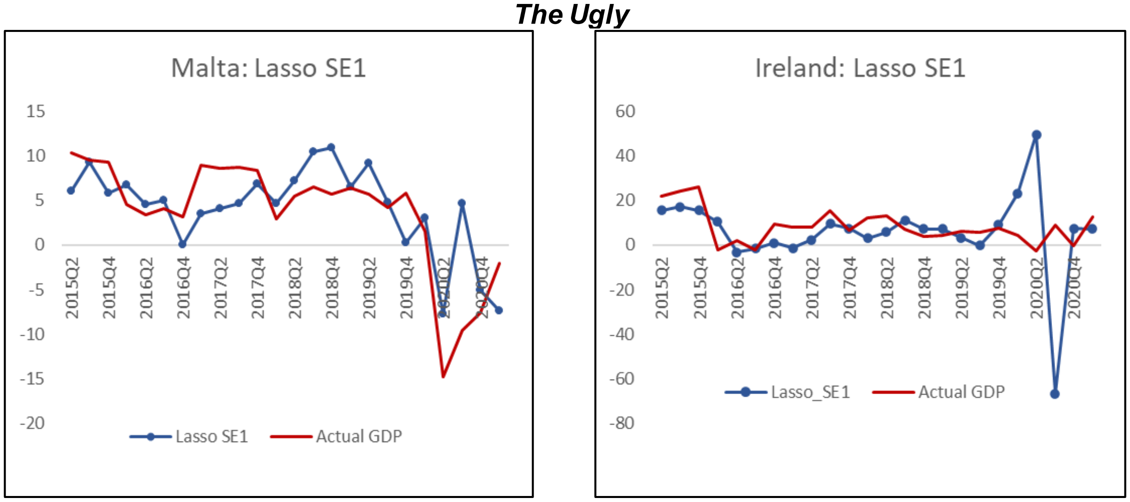


Table 2. Full Sample- Model Performance (RMSE) Until 2021Q1

Country	Austria	Hungary	Ireland	Malta	Poland	Portugal
Backtest period	2017Q1-2021Q1	2016Q1-2021Q1	2015Q2-2021Q1	2012Q-2021Q1	2015Q2-2021Q1	2015Q3-2021Q1
Benchmark- RMSE						
AR1	3.7	3.8	7.7	3.57	2.1	4.0
Machine Learning Models-RMSE						
ElNet_SEm	1.29	1.22	20.34	3.67	1.64	1.18
ElNet_SE1	1.27	1.27	20.34	3.62	1.61	1.18
Lasso_SEm	1.33	1.21	20.13	3.61	1.71	1.20
Lasso_SE1	1.30	1.30	20.11	3.59	1.67	**1.17**
Ridge_SEm	1.26	**1.20**	20.44	3.74	1.45	1.25
Ridge_SE1	**1.23**	1.24	20.44	3.63	**1.45**	1.25
RFor	3.03	3.16	7.86	4.35	1.80	3.02
SVMachine	1.65	1.39	15.87	4.14	1.60	1.62
LinReg	1.38	1.56	14.44	5.57	2.20	1.53
GPReg	2.76	2.71	**6.62**	4.19	1.53	3.62
NN	1.82	2.07	8.09	3.94	1.98	3.05
NNdeep	2.40	2.02	7.57	3.32	2.25	2.56
NNlstm	3.01	2.57	8.39	5.06	2.22	2.69
NNbilstm	2.78	2.13	7.59	4.55	2.45	2.87
NNconv	2.11	2.22	10.46	4.50	1.96	1.71
Dynamic Factor Model-RMSE						
DFM	1.88	2.11	7.13	**2.77**	1.71	2.80

Source: IMF staff calculation.

Note: the green highlighted cells indicate models with lower RMSE against the AR1 benchmark, while the bold red values indicate a model with the lowest RMSE for each country.

Table 3. Pre-COVID Samples- Model Performance (RMSE) Until 2019Q4

Country	Austria	Hungary	Ireland	Malta	Poland	Portugal
Backtest period	2017Q1-2019Q4	2016Q1-2019Q4	2015Q2-2019Q4	2012Q-2019Q4	2015Q2-2019Q4	2015Q3-2019Q4
Benchmark- RMSE						
AR1	0.57	0.9	7.6	**2.11**	0.7	0.5
Machine Learning Models-RMSE						
ElNet_SEm	0.57	1.13	6.28	3.15	0.70	0.42
ElNet_SE1	0.61	1.14	6.28	3.14	0.70	0.43
Lasso_SEm	0.60	1.12	6.53	3.07	0.66	0.41
Lasso_SE1	0.60	1.19	6.57	3.05	**0.66**	0.41
Ridge_SEm	0.58	1.11	6.24	3.14	0.71	0.44
Ridge_SE1	0.60	1.11	6.24	3.14	0.71	0.44
RFor	0.69	0.83	8.46	4.13	1.01	0.93
SVMachine	0.65	1.30	6.19	3.03	0.77	0.43
LinReg	0.60	1.36	7.80	3.52	0.66	0.41
GPReg	0.75	0.71	**4.79**	2.70	0.67	**0.35**
NN	0.90	0.94	6.33	3.40	0.81	0.76
NNdeep	1.02	1.26	6.19	2.68	0.83	0.66
NNlstm	0.54	0.78	8.72	4.09	1.23	0.47
NNbilstm	1.01	0.77	7.26	3.02	1.15	0.51
NNconv	0.90	1.34	6.68	2.69	0.92	0.45
Dynamic Factor Model-RMSE						
DFM	**0.47**	**0.60**	7.43	2.36	0.92	0.46

Source: IMF staff calculation.

Note: the green highlighted cells indicate models with lower RMSE against the AR1 benchmark, while the bold red values indicate a model with the lowest RMSE for each country.

Table 4. During COVID-19 Sample- Model Performance (RMSE) Between 2020Q1-2021Q1

Country	Austria	Hungary	Ireland	Malta	Poland	Portugal
Backtest period	2020Q1-2021Q1	2020Q1-2021Q1	2020Q1-2021Q1	2020Q1-2021Q1	2020Q1-2021Q1	2020Q1-2021Q1
Benchmark- RMSE						
AR1	6.86	7.6	8.2	8.0	5.1	8.4
Machine Learning Models-RMSE						
ElNet_SEm	2.21	1.49	42.84	7.81	2.85	2.40
ElNet_SE1	2.14	1.60	42.84	7.65	2.85	2.40
Lasso_SEm	2.26	1.47	42.23	7.22	2.96	2.47
Lasso_SE1	2.20	1.60	42.16	7.62	3.06	**2.40**
Ridge_SEm	2.15	**1.42**	43.09	7.84	2.84	2.55
Ridge_SE1	**2.06**	1.56	43.09	7.80	2.80	2.55
RFor	5.49	6.30	**4.97**	9.98	5.54	6.23
SVMachine	2.88	1.65	32.61	7.38	**2.75**	3.39
LinReg	2.38	2.08	27.74	12.10	4.32	3.18
GPReg	4.95	5.41	11.10	11.13	5.26	7.74
NN	3.05	3.89	12.74	12.73	5.02	6.38
NNdeep	4.12	3.47	11.40	8.38	6.15	5.34
NNlstm	5.48	5.08	6.97	9.73	5.30	5.70
NNbilstm	4.87	4.15	8.71	10.13	5.97	6.07
NNconv	3.61	3.86	18.85	7.54	5.46	3.57
Dynamic Factor Model-RMSE						
DFM	3.39	4.20	5.87	**4.85**	4.02	5.94

Source: IMF staff calculation.

Note: the green highlighted cells indicate models with lower RMSE against the AR1 benchmark, while the bold red values indicate a model with the lowest RMSE for each country.

To operationalize our findings, we developed an integrated, fully-automated nowcasting tool. The tool automatically (i) collects and treats the dataset; (ii) applies a suit of DFM and ML models to the dataset to perform backtest (charts and indicators of forecasting accuracy); (iii) re-estimate the model each time new data becomes available and produces a nowcast of the current quarter GDP growth; and, (iv) generates an aggregated output for all the methods across the whole period and the subsamples. Once variables to form the dataset have been selected, the tool can be easily applied to any country, subject to data availability (Annex [2]).

Summary and Conclusion

The COVID-19 pandemic has underscored the need for timely data and methods, which allow for the assessment of economies' current health. We motivate, compile, and discuss datasets for six European economies and apply and compare traditional and new nowcasting methods for the period 1995Q1-2021Q4 as well as during normal times (1995-2019) and during the COVID-19 pandemic (2020Q1 up to 2021q4). The datasets comprise standard variables suggested by the literature as well as new variables that have become recently available such as Google Trends and air quality. We provide lists of variables, and their transformations, for six European countries that can serve as examples for other economies. To avoid issues linked to large datasets (collinearity between explanatory variables and model overfitting), we apply methods capable of reducing the dimension of the original dataset. As a representative of a standard approach, we apply a DFM following the New York Fed tradition. Building on the growing evidence of the usefulness of machine learning methods in economics, we introduce several ML algorithms to our exercise.

Overall, we find that the tools we applied add value and have the power to inform the nowcast of the current quarter GDP growth. Applying alternative methods and comparing their predictive capability on historical data, we find that a strong majority of our methods outperform the AR(1) benchmark model. Specifically, applying the ML methods reduced the average forecast errors up to 75 percent and the DFM reduced the forecast error by half compared to the AR(1) model across all countries. Nonetheless, we found that there is no *one-size-fits-all* method that would outperform the remaining methods in case of all countries under all circumstances. Although our results apply to a specific set of countries and time episodes—and cannot necessarily be generalized—we find that our DFM models tend to perform better during normal times but not as much during times of heightened volatility. In other words, DFMs did not predict the sharp contraction of 2020Q2 to the extent as the ML methods we used did. On the contrary, most of the ML methods performed strongly at identifying turning points, while it seems to overvalue the effect of minor changes in some of the explanatory variables on GDP growth during normal times. Therefore, when using the techniques we applied, the practitioner needs to carefully analyze back-testing results and apply judgment as to what models can be trusted to provide the most informative nowcast.

Appendices

1. Nowcasting Variables by Country

Austria

Indicator	Global	Real	Financial	External	Frequency	Transformation
National accounts						
Real gross domestic product	✖	✖			Q	YoY percent change
Real final consumption expenditure	✖	✖			Q	YoY percent change
Real gross capital formation	✖	✖			Q	YoY percent change
Real exports of goods & services	✖	✖			Q	YoY percent change
Real imports of goods & services	✖	✖			Q	YoY percent change
Housing and constrution						
Residential property prices	✖	✖			Q	Level
Industrial production index: construction	✖	✖			M	YoY percent change
Labor						
Labor force participation rate	✖	✖			Q	Level
Employment rate	✖	✖			Q	Level
Unemployment rate	✖	✖			Q	Level
Productivity per employee: industry incl. construction	✖	✖			M	YoY percent change
Manufacturing						
Domestic industrial new orders	✖	✖			M	YoY percent change
Industrial production index: total industry excl. construction	✖	✖			M	YoY percent change
Germany: industrial production excl. construction	✖	✖		✖	M	YoY percent change
Germany: capacity utilization: manufacturing	✖			✖	Q	Level
Germany: volume of manufacturing orders	✖			✖	M	YoY percent change
EU27: industrial production: industry incl. construction	✖			✖	M	YoY percent change
United States: industrial production index	✖			✖	M	YoY percent change
Euro area: manufacturing new orders	✖			✖	M	Level
Retail and consumption						
Retail trade value excl. autos & motorcycles	✖	✖			M	YoY percent change
Retail trade volume excl. autos & motorcycles	✖	✖			M	YoY percent change
New passenger car registrations	✖	✖			M	YoY percent change
International trade						
Merchandise trade: exports	✖	✖			M	YoY percent change
Merchandise trade: imports	✖	✖			M	YoY percent change
Balance of payments: services credit	✖	✖			M	YoY percent change
Balance of payments: services debit	✖	✖			M	YoY percent change
Financial						
Domestic loans to consumers	✖	✖	✖		M	YoY percent change
Households' overnight deposits	✖	✖	✖		M	YoY percent change
Surveys						
Industrial confidence indicator	✖	✖			M	Level
Purchasing Managers Index: Manufacturing	✖	✖			M	Level
Germany: Ifo business climate	✖			✖	M	Level
Euro area: Purchasing Managers Index: composite employment	✖			✖	M	Level
Euro area: business climate indicator	✖			✖	M	Level
Other						
Tourist arrivals	✖	✖			M	YoY percent change
Tourists' overnight stays	✖	✖			M	YoY percent change
Google: keywords	✖	✖			M	Level
Nitrogen dioxide emissions	✖	✖			M	Level
Apple Mobility Trends Index: walking	✖	✖			D	Level
Google: change in visits relative to baseline: retail & recreation	✖	✖			D	Level

Hungary

Indicator	Global	Real	Financial	External	Frequency	Transformation
National accounts						
Real gross domestic product	✖	✖			Q	YoY percent change
Real final consumption expenditure	✖	✖			Q	YoY percent change
Real gross capital formation	✖	✖			Q	YoY percent change
Real exports of goods & services	✖	✖			Q	YoY percent change
Real imports of goods & services	✖	✖			Q	YoY percent change
Labor						
Unemployment rate, 25-74 years	✖	✖			M	Level
Manufacturing						
Domestic turnover: manufacturing	✖			✖	M	Level
Domestic turnover: consumer goods	✖			✖	M	Level
Domestic turnover: capital goods	✖			✖	M	YoY percent change
Domestic turnover: nondurable consumer goods	✖			✖	M	YoY percent change
Domestic turnover: durable consumer goods	✖			✖	M	YoY percent change
Domestic turnover: intermediate goods	✖			✖	M	Level
Retail and consumption						
New car registrations	✖	✖			M	Level
Financial						
Stock price index, BUX	✖	✖			M	Level
Surveys						
Business confidence index	✖	✖			M	Level
Economic sentiment index	✖	✖			M	YoY percent change
Consumer confidence index	✖	✖			M	YoY percent change
Retail trade survey: volume of stocks	✖	✖			M	YoY percent change
Retail trade survey: present business situation	✖	✖			M	YoY percent change
Retail trade survey: orders placed with suppliers	✖	✖			M	YoY percent change
Retail trade survey: expected business situation	✖	✖			M	YoY percent change
Consumer survey: financial situation, next 12 months	✖	✖			M	YoY percent change
Consumer survey: general economic situation, next 12 months	✖	✖			M	YoY percent change
Consumer survey: major purchases, next 12 months	✖	✖			M	YoY percent change
Consumer survey: major purchases, at present	✖	✖			M	YoY percent change
Consumer survey: price trends, next 12 months	✖	✖			M	YoY percent change
Consumer survey: savings, next 12 months	✖	✖			M	YoY percent change
Manufacturing PMI	✖	✖			M	YoY percent change
Manufacturing PMI: output	✖	✖			M	YoY percent change
Manufacturing PMI: quantity of purchases	✖	✖			M	YoY percent change
Manufacturing PMI: stock of finished goods	✖	✖			M	YoY percent change
Manufacturing PMI: employment	✖	✖			M	YoY percent change
Manufacturing PMI: imports	✖	✖			M	YoY percent change
Manufacturing PMI: stock of purchases	✖	✖			M	Level
Manufacturing PMI: new orders	✖	✖			M	Level
Manufacturing PMI: input price	✖	✖			M	YoY percent change
Manufacturing PMI: supplier deliveries	✖	✖			M	YoY percent change
Manufacturing PMI: exports	✖	✖	✖		M	YoY percent change
Factors limiting building activity: demand	✖	✖	✖		M	YoY percent change
Factors limiting building activity: financial constraints	✖	✖	✖		M	YoY percent change
Factors limiting building activity: labor shortage	✖	✖			M	YoY percent change
Factors limiting building activity: equipment shortage	✖	✖			M	YoY percent change
Factors limiting building activity: other factos	✖	✖		✖	M	YoY percent change
Industry: volume of export order books	✖			✖	M	Level
Industry: production expectations	✖			✖	M	Level
Industry: production trend in recent months	✖	✖			M	Level
Industry: stocks of finished products	✖	✖			M	Level
Services: expected demand, next 3 months	✖	✖			M	Level
Services: price expectations, next 3 months	✖	✖			M	Level
Other						
Google: keywords	✖	✖			M	Level
Nitrogen dioxide emissions	✖	✖			M	Level

Ireland

Indicator	Global	Soft	Real	Financial	External	Frequency	Transformation
National accounts							
Real gross domestic product	✖		✖			Q	YoY percent change
Real private consumption expenditure	✖		✖			Q	YoY percent change
Real gross fixed capital formation	✖		✖			Q	YoY percent change
Real gross fixed capital formation, building and construction	✖		✖			Q	YoY percent change
Real gross fixed capital formation, machinery and equipment	✖		✖			Q	YoY percent change
Real gross fixed capital formation, intabgible assets	✖		✖			Q	YoY percent change
Real exports of goods & services	✖		✖			Q	YoY percent change
Real imports of goods & services	✖		✖			Q	YoY percent change
Real government consumption expenditure	✖		✖			Q	YoY percent change
Real modified total domestic demand	✖		✖			Q	YoY percent change
Real gross value added (GVA)	✖		✖			Q	YoY percent change
Real GVA: agriculture, forestry, and fisheries	✖		✖			Q	YoY percent change
Real GVA: industry excluding construction	✖		✖			Q	YoY percent change
Real GVA: manufacturing	✖		✖			Q	YoY percent change
Real GVA: building and construction	✖		✖			Q	YoY percent change
Real GVA: distribution, transport, hotels, and restaurants	✖		✖			Q	YoY percent change
Real GVA: information and communication	✖		✖			Q	YoY percent change
Real GVA: financial and insurance activities	✖		✖			Q	YoY percent change
Real GVA: real estate activities	✖		✖			Q	YoY percent change
Real GVA: professional, admin, and support services	✖		✖			Q	YoY percent change
Real GVA: public admin, education, and health	✖		✖			Q	YoY percent change
Real GVA: arts, entertainment, and other services	✖		✖			Q	YoY percent change
Housing and construction							
Industrial production volume: building and construction	✖		✖			Q	YoY percent change
Prices							
Harmonized inflation	✖					M	YoY percent change
HICP: Total excluding energy, food, alcohol, and tobacco	✖					M	YoY percent change
Residential property price index	✖		✖			M	YoY percent change
Industrial price index: transportable goods industries	✖					M	YoY percent change
West Texas intermediate crude oil spot price (Cushing, OK)	✖				✖	M	Percent change
Exchange rate: US dollar/euro (average)	✖				✖	M	Percent change
Exchange rate: UK Pound sterling/euro (average)	✖				✖	M	Percent change
Labor							
Employed	✖		✖			Q	YoY percent change
Unemployment rate, Covid-19 adjusted, upper bound	✖		✖			M	Level
Unemployment rate, Covid-19 adjusted, lower bound	✖		✖			M	Level
Unemployment	✖		✖			M	Percent change
Pandemic unemployment payment recipients	✖		✖			M	Level
Participation rate	✖		✖			Q	YoY percent change
Average hourly earnings	✖		✖			Q	YoY percent change
Public sector employment including semi-state bodies	✖		✖			Q	YoY percent change
Total job postings	✖		✖			D	Level

Indicator	Global	Soft	Real	Financial	External	Frequency	Transformation
Fiscal							
Central government: tax revenue	✖		✖			M	YoY percent change
Central government: tax revenue, income tax	✖		✖			M	YoY percent change
Central government: tax revenue, value added tax	✖		✖			M	YoY percent change
Central government: tax revenue, corporation tax	✖		✖			M	YoY percent change
Central government: tax revenue, excise duty tax	✖		✖			M	YoY percent change
Manufacturing							
Industrial production	✖		✖			M	Percent change
Industrial production: modern sector	✖		✖			M	YoY percent change
Industrial production: traditional sector	✖		✖			M	YoY percent change
Euro area: manufacturing new orders	✖				✖	M	Percent change
EU27: industrial production: industry incl. construction	✖				✖	M	Percent change
United States: industrial production index	✖				✖	M	Percent change
Retail and consumption							
Services value index	✖		✖			M	Percent change
Electric power average actual load per 30 minutes	✖		✖			D	YoY percent change
Retail sales value	✖		✖			M	Percent change
Gross new spending on credit and debit cards	✖		✖			D	Level
International trade							
Merchandise trade: exports	✖		✖			M	Percent change
Merchandise trade: imports	✖		✖			M	Percent change
Financial							
Stock price index, ISEQ	✖			✖		M	Percent change
3-month interbank market rate	✖			✖		M	Level
10-year government bond yield	✖			✖		M	Level
Money supply: M3, growth rate	✖			✖		M	Level
Private sector loan growth	✖			✖		M	Level
Stock price index, S&P 500	✖			✖	✖	M	Percent change
S&P Goldman Sachs commodity index: energy	✖			✖	✖	M	Percent change
S&P Goldman Sachs commodity index: non-energy	✖			✖	✖	M	Percent change
Stock price index, FTSE Eurofirst 80	✖			✖	✖	M	Percent change
Euro area: deposit rate (end of period)	✖			✖	✖	M	Level
Surveys							
Purchasing Managers Index (PMI): composite output	✖	✖				M	Percent change
PMI: manufacturing	✖	✖				M	Percent change
PMI: construction	✖	✖				M	Percent change
Consumer sentiment index, 3-month moving average	✖	✖				M	Change (difference)
Consumer sentiment index: consumer expectations, 3-month moving average	✖	✖				M	Change (difference)
Euro area PMI: composite output	✖	✖			✖	M	Percent change
Euro area: business climate indicator	✖				✖	M	Change (difference)
Euro area: consumer confidence indicator	✖				✖	M	Change (difference)
Other							
Apple Mobility Trends Index: walking	✖		✖			D	Level
Moovit: public transit app usage index	✖		✖			D	Level
Google: change in visits relative to baseline: retail & recreation	✖		✖			D	Level
Composite indicator of systemic stress	✖		✖			D	Level
Google: keywords	✖	✖				M	Level
Nitrogen dioxide emissions	✖	✖				M	Level

Malta

Indicator	Global	Soft	Real	Financial	External	Frequency	Transformation
National accounts							
Real gross domestic product	✖		✖			Q	YoY percent change
Real private consumption expenditure	✖		✖			Q	YoY percent change
Real government consumption expenditure	✖		✖			Q	YoY percent change
Real gross capital formation	✖		✖			Q	Level
Real exports of goods & services	✖		✖			Q	Level
Real imports of goods & services	✖		✖			Q	Level
Housing and constrution							
Building permits: residential buildings excl. community residences	✖		✖			Q	Level
Construction output	✖		✖		✖	Q	Level
Labor							
Employment rate	✖		✖			Q	Level
Unemployment rate, NSA	✖		✖			Q	YoY percent change
Hours worked: industry excl. construction	✖		✖			Q	YoY percent change
Hours worked: construction	✖		✖			Q	Level
Unemployment rate, SA	✖		✖			Q	Level
Fiscal							
Tax revenue	✖		✖			M	YoY percent change
Manufacturing							
Industrial turnover	✖		✖			Q	YoY percent change
Industrial production	✖		✖			Q	YoY percent change
EU27: industrial production: industry incl. construction	✖				✖	Q	YoY percent change
EU27: industry: capacity utilization	✖		✖		✖	Q	YoY percent change
Retail and consumption							
Retail trade: confidence indicator	✖	✖				Q	YoY percent change
Wholesale & retail trade & repair of autos & motorcycles, volume	✖		✖			Q	YoY percent change
International trade							
Merchandise trade: exports	✖		✖			Q	YoY percent change
Merchandise trade: imports	✖		✖			Q	YoY percent change
Services trade, value	✖		✖			Q	YoY percent change
Financial							
Other MFI loans to households & individuals for house purchase	✖			✖		Q	Hp-filtered cyclical components
Term premium	✖			✖		M	YoY percent change
Surveys							
Industrial confidence indicator	✖					M	YoY percent change
Services confidence indicator	✖	✖				M	YoY change (difference)
Consumer confidence indicator	✖	✖				M	YoY percent change
Economic sentiment indicator	✖	✖				M	YoY percent change
Retail trade: order expectations	✖	✖				M	YoY percent change
Services: expected demand, next 3 months	✖	✖				M	YoY percent change
Construction: employment expectations, next 3 months	✖	✖				M	YoY change (difference)
Industry: volume of order books	✖	✖				Q	YoY change (difference)
Industry: volume of export order books	✖	✖				M	YoY percent change
Euro area: economic sentiment indicator	✖	✖			✖	M	YoY percent change
Euro area: manufacturing new orders	✖	✖			✖	M	YoY percent change
Euro area: business climate indicator	✖	✖			✖	M	YoY change (difference)
Other							
Departing tourists	✖		✖		✖	Q	YoY percent change
Departing tourists: total expenditure	✖		✖			Q	YoY percent change
Departing tourists: total nights	✖		✖		✖	Q	YoY percent change
Stock of licensed motor vehicles	✖		✖			D	YoY percent change
Google: keywords	✖		✖			M	YoY change (difference)
Nitrogen dioxide emissions	✖		✖			M	YoY change (difference)

Poland

Indicator	Global	Soft	Real	Labor	Financial	External	Frequency	Transformation
National accounts								
Real gross domestic product	✖		✖				Q	YoY percent change
Real final consumption expenditure	✖		✖				Q	YoY percent change
Real government consumption expenditure	✖		✖				Q	YoY percent change
Real private consumption expenditure	✖		✖				Q	YoY percent change
Real gross capital formation	✖		✖				Q	YoY percent change
Real gross fixed capital formation	✖		✖				Q	YoY percent change
Real exports of goods & services	✖		✖				Q	YoY percent change
Real imports of goods & services	✖		✖				Q	YoY percent change
EU27: real gross domestic product	✖					✖	Q	YoY percent change
United States: real gross domestic product	✖					✖	Q	YoY percent change
United States: IHS Markit monthly real GDP	✖					✖	M	YoY percent change
Housing and constrution								
Construction & assembly production	✖		✖				M	YoY percent change
Prices								
HICP: core inflation	✖						M	YoY percent change
West Texas intermediate crude oil spot price (Cushing, OK)	✖				✖	✖	M	YoY percent change
Poland: JPMorgan Real Broad Effective Exchange Rate Index, CPI I	✖				✖	✖	M	Level
Labor								
Unemployment rate	✖			✖			M	Level
Average paid employment: enterprise sector	✖			✖			M	YoY percent change
Germany: unemployment rate	✖					✖	M	Level
Germany: output per working hour, industry incl. construction	✖					✖	M	YoY percent change
Manufacturing								
EU27: industrial production: industry incl. construction	✖					✖	M	YoY percent change
Industrial new orders	✖		✖				M	YoY percent change
United States: industrial production	✖					✖	M	YoY percent change
Industrial production excl. construction	✖		✖				M	YoY percent change
Industrial production: mining & quarrying	✖		✖				M	YoY percent change
Industrial production: manufacturing	✖		✖				M	YoY percent change
Germany: industrial production incl. construction	✖					✖	M	YoY percent change
Germany: industrial production, construction	✖					✖	M	YoY percent change
Germany: manufacturing orders, volume	✖					✖	M	YoY percent change
Retail and consumption								
Wholesale trade enterprises	✖		✖				M	YoY percent change
Retail sales, constant prices	✖		✖				M	YoY percent change
International trade								
Balance of payments: exports	✖		✖				M	YoY percent change
Balance of payments: imports	✖		✖				M	YoY percent change
Financial								
MSCI share price index with net dividends, local currency	✖				✖		M	YoY percent change
Money supply: M3	✖				✖		M	YoY percent change
Poland: policy rate	✖				✖		M	Level
Stock price index, FTSE Eurofirst 80	✖				✖	✖	M	YoY percent change
Surveys								
PMI: manufacturing	✖						M	Level
Manufacturing index of overall economic climate	✖	✖					M	Level
Euro area: business climate indicator	✖	✖				✖	M	Level
Euro area: consumer confidence indicator	✖	✖				✖	M	Level
Euro area: sentix overall index	✖	✖				✖	M	Level
Euro area PMI: manufacturing new orders, domestic market	✖	✖				✖	M	Level
Euro area PMI: composite output	✖					✖	M	Level
Global PMI: services	✖					✖	M	Level
Developed markets PMI: services	✖					✖	M	Level
Europe PMI: services	✖					✖	M	Level
United States PMI: services	✖					✖	M	Level
Germany: ifo business climate index, industry & trade	✖					✖	M	Level
Germany: ifo business expectations, industry & trade	✖					✖	M	Level
Germany: ifo production activity, manufacturing incl. food & beve	✖					✖	M	YoY percent change
Germany: ifo export expectations next 3 months, manufacturing ir	✖					✖	M	Level
Germany: sentix overall economic index	✖					✖	M	Level
Germany: IAB labor market barometer	✖					✖	M	YoY percent change
Other								
Apple Mobility Trends Index: walking	✖	✖	✖	✖	✖	✖	D	Level
Google: change in visits relative to baseline: transit stations	✖	✖	✖	✖	✖	✖	D	Level
Google: change in visits relative to baseline: retail & recreation	✖	✖	✖	✖	✖	✖	D	Level
Tourist arrivals	✖		✖				M	YoY percent change
Google: keywords	✖		✖				M	YoY percent change
Nitrogen dioxide emissions	✖		✖				M	YoY percent change

Portugal

Indicator	Global	Real	Financial	External	Frequency	Transformation
National accounts						
Real gross domestic product	✖	✖			Q	YoY percent change
Real final consumption expenditure	✖	✖			Q	YoY percent change
Real government consumption expenditure	✖	✖			Q	YoY percent change
Real exports of goods & services	✖	✖			Q	YoY percent change
Real imports of goods & services	✖	✖			Q	YoY percent change
Gross value added: construction	✖	✖			Q	YoY percent change
Housing and constrution						
Housing price index	✖	✖			Q	Level
Construction output	✖	✖			M	YoY percent change
Labor						
Participation rate	✖	✖			Q	YoY percent change
Employment rate	✖	✖			Q	YoY percent change
Unemployment rate	✖	✖			Q	YoY percent change
Gross monthly earnings, 3-month moving average	✖	✖			M	YoY percent change
Real labor productivity per person	✖	✖			Q	YoY percent change
Manufacturing						
Industrial new orders	✖	✖			M	YoY percent change
Industrial production	✖	✖			M	YoY percent change
Germany: industrial production excl. construction	✖	✖		✖	M	YoY percent change
Germany: capacity utilization, manufcturing	✖			✖	Q	Level
Germany: manufacturing orders, volume	✖			✖	M	YoY percent change
EU27: industrial production: industry incl. construction	✖			✖	M	YoY percent change
United States: industrial production index	✖			✖	M	YoY percent change
Euro area: manufacturing new orders	✖			✖	M	Level
Retail and consumption						
Retail trade	✖	✖			M	YoY percent change
Retail sales volume	✖	✖			M	YoY percent change
Total industry sales	✖	✖			M	YoY percent change
New car registrations	✖	✖			M	YoY percent change
International trade						
Merchandise trade: exports	✖	✖			M	YoY percent change
Merchandise trade: imports	✖	✖			M	YoY percent change
Services exports	✖	✖			M	YoY percent change
Services imports	✖	✖			M	YoY percent change
Services trade value	✖	✖			M	YoY percent change
Financial						
OMFI loans to private individuals for consumption	✖	✖	✖		M	YoY percent change
Overnight deposits	✖	✖	✖		M	YoY percent change
Surveys						
Economic sentiment indicator	✖	✖			M	Level
Manufacturing confidence	✖	✖			M	Level
Germany: ifo business climate	✖			✖	M	Level
Euro area PMI: composite employment	✖			✖	M	Level
Euro area: business climate indicator	✖			✖	M	Level
Other						
Number of travelers	✖	✖			M	YoY percent change
Overnight stays in tourist accommodation establishments	✖	✖			M	YoY percent change
Oxford government response index	✖	✖			D	Level
Oxford stringency index	✖	✖			D	Level
Oxford risk of openness index: community understanding	✖	✖			D	Level
Google: keywords	✖	✖			M	Level
Nitrogen dioxide emissions	✖	✖			M	Level

2. Sample Results—Hungary

This appendix presents an example of a standardized set of results generated by the automated tool. These specific results refer to Hungary as of November 2011.

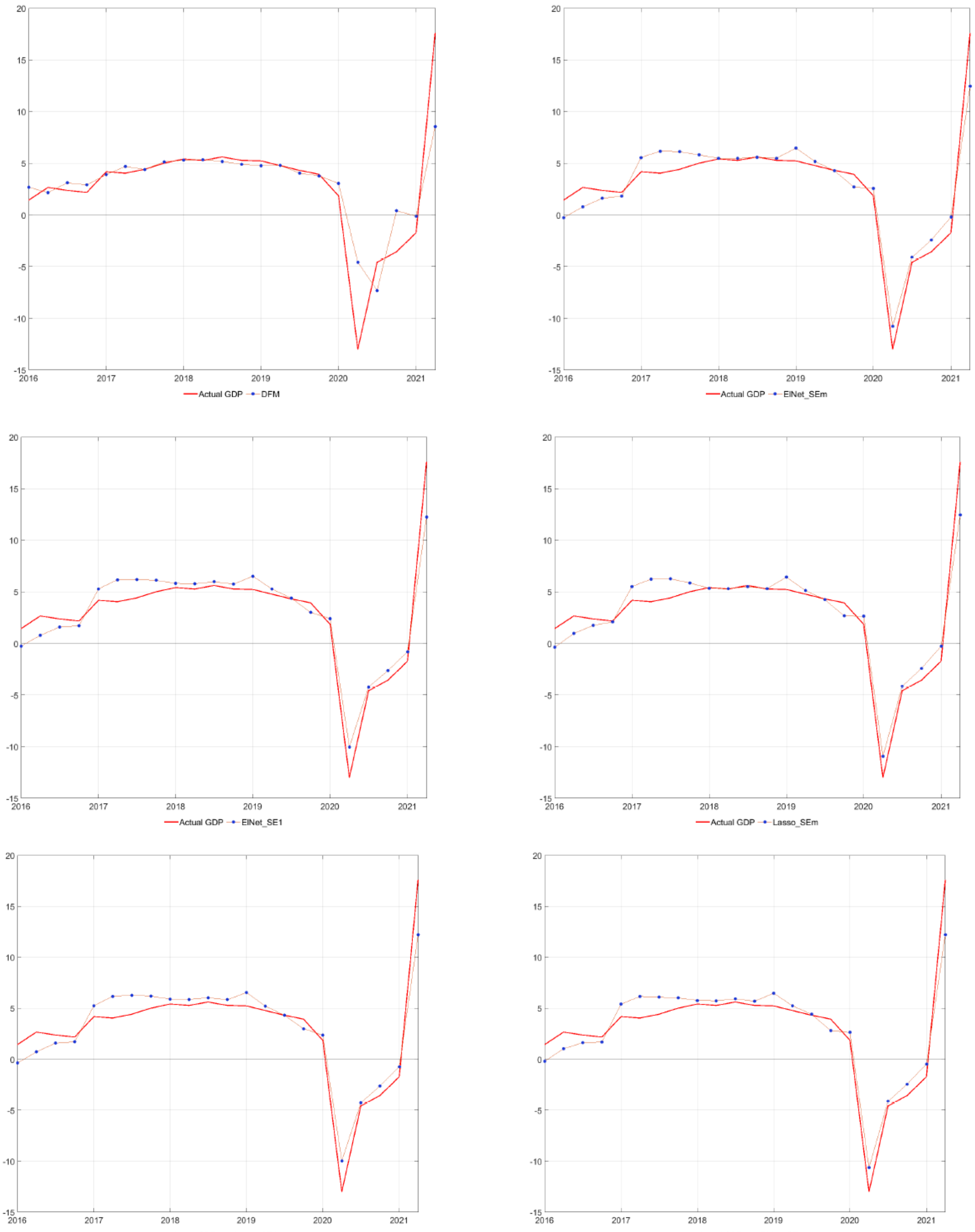
The first part (Backtest) provides indicators (MFE, RMSE, and MDA) allowing for an assessment of the forecasting power of individual models included in the tool on historical data over three different periods (2016Q1 to 2021Q2; 2016Q1 to 2019Q4; and 2020Q1 to 2021Q2). The charts compare the actual GDP growth with out-of-sample prediction by individual models. The indicators and the charts should provide some guidance on the accuracy and reliability of individual models.

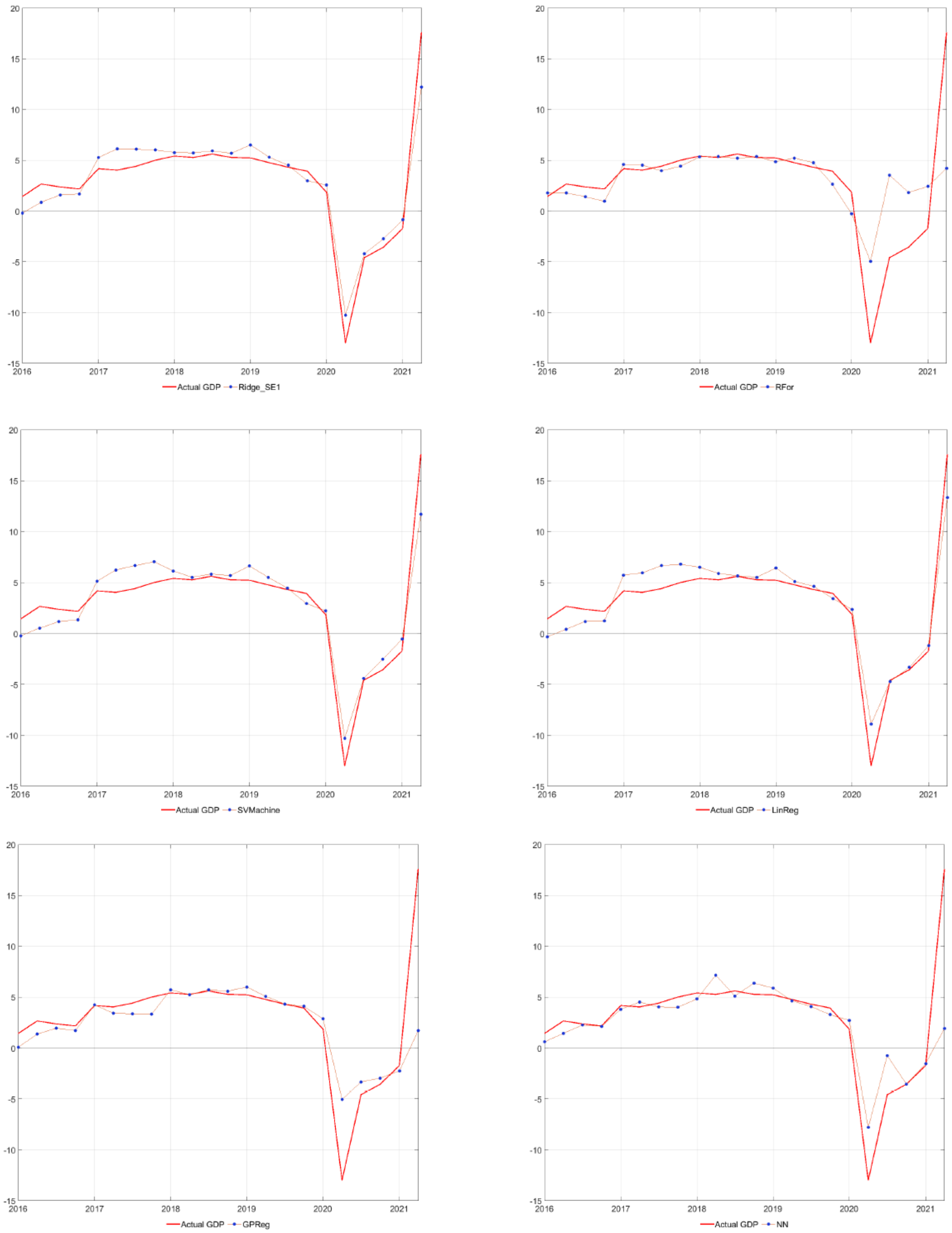
The second part (Nowcasting Tracker) presents current quarter GDP growth as predicted by individual models. The table presents results for individual models ranked by the size of their RMSE quantified on historical data, e.g., lower RMSE indicates better predictive accuracy of a model on historical data. Finally, the nowcasting chart presents historical GDP growth and the values expected for the current quarter by the six most accurate models.

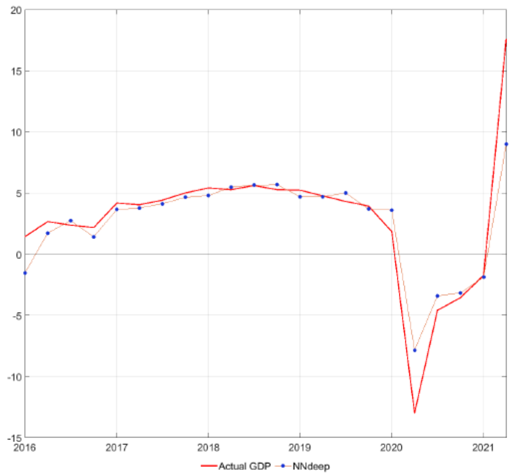
Nowcasting models: Backtest

	2016Q1 to 2021Q2			2016Q1 to 2019Q4			2020Q1 to 2021Q2		
	MFE	RMSE	MDA	MFE	RMSE	MDA	MFE	RMSE	MDA
DFM	-0.202	2.893	0.81	-0.063	0.507	0.733	-0.574	5.477	1
ElNet_SEm	-0.14	1.6	0.81	-0.133	1.132	0.733	-0.16	2.444	1
ElNet_SE1	-0.196	1.662	0.714	-0.248	1.146	0.6	-0.059	2.575	1
Lasso_SEm	-0.129	1.591	0.857	-0.132	1.134	0.8	-0.121	2.419	1
Lasso_SE1	-0.201	1.691	0.762	-0.255	1.183	0.667	-0.059	2.599	1
Ridge_SEm	-0.194	1.621	0.714	-0.231	1.109	0.6	-0.097	2.519	1
Ridge_SE1	-0.176	1.632	0.714	-0.234	1.108	0.6	-0.021	2.548	1
RFor	-0.282	4.089	0.762	0.248	0.65	0.667	-1.695	7.758	1
SVMachine	-0.186	1.827	0.714	-0.284	1.342	0.6	0.075	2.728	1
LinReg	-0.262	1.707	0.81	-0.297	1.333	0.733	-0.168	2.438	1
GPReg	0.477	3.866	0.81	0.305	0.759	0.733	0.934	7.298	1
NN	0.341	3.685	0.619	0.119	0.792	0.467	0.931	6.937	1
NNdeep	0.281	2.315	0.762	0.368	0.881	0.667	0.047	4.194	1
NNlstm	-0.07	3.351	0.714	0.174	0.736	0.6	-0.721	6.304	1
NNbilstm	-0.081	4.024	0.571	0.269	0.989	0.4	-1.015	7.535	1
NNconv	-0.363	2.374	0.762	-0.313	1.544	0.667	-0.499	3.782	1
AR1	0.571	5.374	0.714	0.204	0.862	0.667	1.547	10.193	0.8

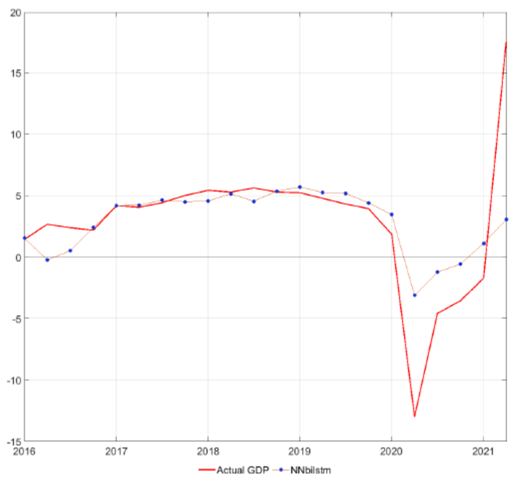
Note: MFE … Mean Forecast Error, RMSE … Residual Mean Squared Error, and MDA… Mean Directional Accuracy

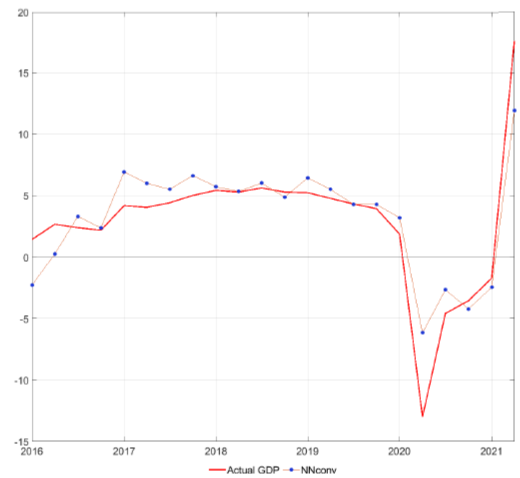

20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
DFM
ElNet_SEm
ElNet_SE1
Lasso_SEm
Lasso_SE1
Ridge_SEm


Actual GDP
Ridge_SE1
Actual GDP
RFor
Actual GDP
SVMachine
Actual GDP
LinReg
Actual GDP
GPReg
Actual GDP
NN


20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
NNdeep


20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
NNlstm

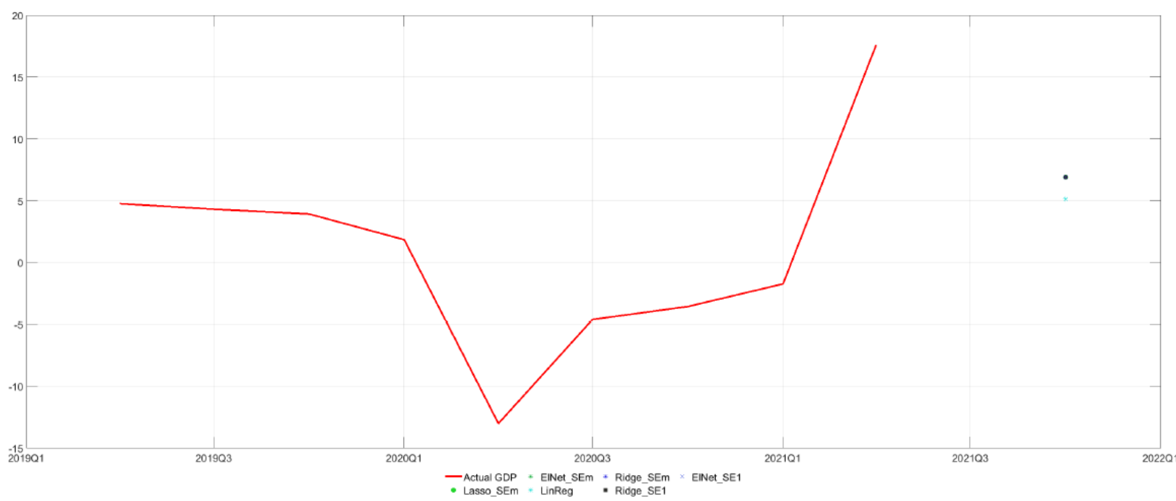

20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
NNbilstm


20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
NNconv


20
15
10
5
0
-5
-10
-15
2016
2017
2018
2019
2020
2021
Actual GDP
AR1

Nowcasting Tracker

	Nowcast	Historical RMSE
Lasso_SEm	6.89	1.69
ElNet_SEm	6.89	1.69
LinReg	5.15	1.70
Ridge_SEm	6.89	1.70
Ridge_SE1	6.89	1.72
ElNet_SE1	6.90	1.75
Lasso_SE1	6.90	1.78
SVMachine	5.63	1.90
NNconv	9.98	2.27
NNdeep	18.65	2.40
DFM	5.17	2.96
NN	33.94	3.26
NNlstm	11.99	3.79
RFor	8.62	3.86
NNbilstm	9.97	3.91
GPReg	10.73	4.02



3. An Integrated Tool

Using the methodologies described in this paper, we developped an integrated tool that automatically (i) collect and treat the data set; (ii) applies a suit of DMF and ML models to the dataset to generate backtest results (graphs and quantitative indicators), and (iii) nowcast GDP growth for the current quarter for each method and aggregating all the methods. The tool can be applied to any country, and is automated once variables to form the dataset have been selected.
The tool is developed in Matlab and is composed of three major components: the Data Warehouse, the Backtest, and the Nowcasting Tracker. All three components are integrated by the Reports Generator, which is a Matlab script that serves as the main user interface (in addition to an Excel-based file that the script reads from.) Normally, it is not necessary for the users to modify or carefully read the Matlab programs associated with the three underlying components, unless they would like to add their own nowcasting method. Even in such cases, the modularity of the toolkit is designed for easy addition of user-defined methods, as we highlight below.

All nowcasting methods, including the DFM and various machine learning algorithms, have their respective standalone functions or scripts and are invoked as needed in the Backtest and Nowcasting Tracker. Thanks to this modularity and standardized output format of the nowcasting methods, with some familiarity to the toolkit structure and the Matlab language, the user can proceed without special assistance to add their customized nowcast methods to the toolkit. Removing methods from the toolkit is even more straightforward.

Following is a description of the design of each of the three components, i.e. the Data Warehouse, the Backtest, and the Nowcasting Tracker.

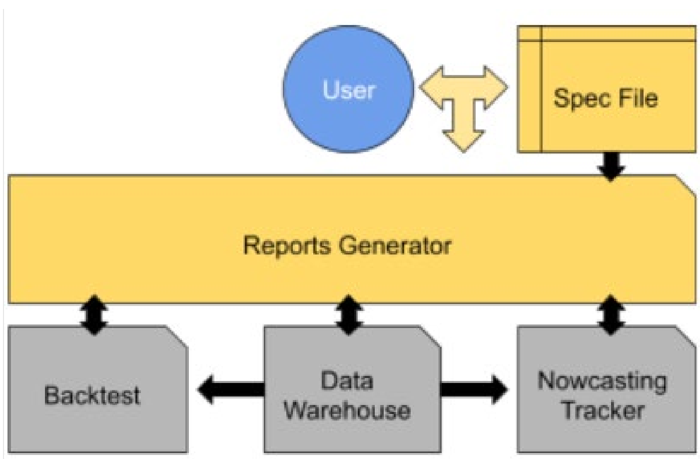


Figure 1

The Data Warehouse is a Matlab script that fetches historical data from various data sources such as Haver and Google. (Where necessary, R scripts are also invoked for retrieving from certain data sources.) The Matlab script receives instructions about what Haver series to fetch, what frequency the data is in, and what data transformation to perform, etc., from an Excel file (the "Spec File,") which is part of the user interface and serves as input to the Reports Generator, as shown in figure 1.

The Backtest performs out-of-sample pseudo-nowcasts on an expanding window of historical data, and produces plots and accuracy metrics for the user to compare the performance of nowcasting methods against one another, as well as against the actual historical outturns. Specifically, out-of-sample pseudo-nowcasts assume that for a certain *historical* quarter t = t', GDP data is available up to t'-1, and try to predict the GDP for quarter t'. The backtest result for a given nowcasting method combines n such out-of-sample nowcasts for t = t', t'-1, …, t'-n+1, where n depends on historical data availability and the user's specification.

Finally, the Nowcasting Tracker produces actual nowcasts for the current quarter using all or some of the methods tested in the backtest. At this point the toolkit has completed a successful execution. A table and a plot of GDP nowcasts are generated and saved for the user's information. The user can base their comprehensive judgement upon each method's backtest performance as well as their knowledge on the economy concerned.

4. Machine Learning Algorithms

1. Regularized Regression methods

Assume a linear model

$$y_t = x_t'\beta + \varepsilon_t$$

where y_t represents a target variable, x_t is a $p \times 1$ vector of features (explanatory variables), and ε_t is the residual.

To estimate a parameter vector β, the *traditional OLS method* minimizes the loss function (residual sum of squares) by solving the quadratic minimization problem:

$$\hat{\beta} \equiv argmin_\beta \left(\|y_t - x_t'\beta\|^2\right)$$

To avoid overfitting in models with a large number of explanatory variables, the *regularized regression methods* penalize the use and size of additional coefficients by expanding the original OLS optimization problem with a penalty term (regularization) into. Depending on the type of penalty, the literature recognizes different types of regularization techniques. Lasso, Ridge and Elastic net are three popular regularization techniques based on the classical linear regression method.

- *Lasso* (Least Absolute Shrinkage and Selection Operator) regression adds absolute value of the coefficients as penalty term to the loss function, the so-called L_1 regularization.

$$\hat{\beta} \equiv argmin_\beta\left(\|y_t - x_t'\beta\|^2 + \lambda_1\|\beta\|\right)$$

- *Ridge regression* modifies the OLS loss function by adding square magnitude of the coefficients as a penalty, the L_2 regularization element,

$$\hat{\beta} \equiv argmin_\beta\left(\|y_t - x_t'\beta\|^2 + \lambda_2\|\beta\|^2\right)$$

- *Elastic net* combines Lasso and Ridge regression by introducing both L_1 and L_2 into the original OLS loss function

$$\hat{\beta} \equiv argmin_\beta\left(\|y_t - x_t'\beta\|^2 + \lambda_1\|\beta\| + \lambda_2\|\beta\|^2\right)$$

Parameters λ_1 and λ_2 (also called hyperparameters) are usually quantified by applying numerical methods (i.e. grid search or random search). The parameters are "tuned" using cross-validation resampling methods by minimizing the forecast error in the validation set.

2. Support Vector Machine

Instead of estimating β by minimizing the sum of squared errors (as in OLS), the Support Vector Machine (SVM) finds β by minimizing the magnitude of the coefficients (L_2 norm) while keeping regression residuals within a specified margin. In other words, SVM method will find an appropriate hyperplane to fit the data while allowing us to define how much error is acceptable in our model.

Similar to linear regression, SVM is also a linear function in order to predict the target variable y_t

$$f(x_t) = x_t'\beta + b$$

Formally, SVM algorithm aims at finding $f(x_t)$ as flat as possible while maintaining the forecast error capped within certain band.[1] Thus, SVM will minimize its loss function by finding a combination of β and observation-specific positive slack constants ξ_t and ξ_t^*:

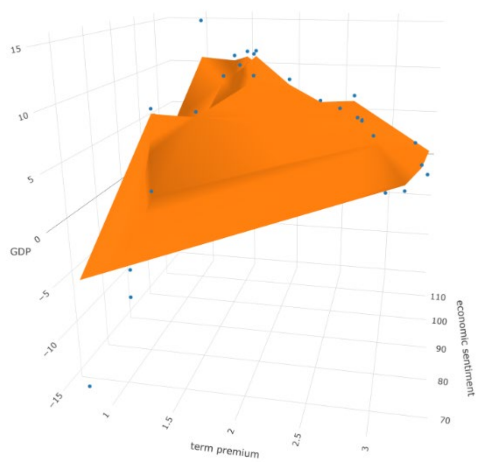
$$\frac{1}{2}\beta'\beta + C\sum_{t=1}^{T}(\xi_t + \xi_t^*)$$

$$s.t. \quad y_t - f(x_t) \leq \varepsilon + \xi_t$$

$$f(x_t) - y_t \leq \varepsilon + \xi_t^*$$

Therefore, at each point, the forecast error is capped by $[-\varepsilon, \varepsilon]$, and can be relaxed by at most ξ_t^* or ξ_t at higher or lower end. The constant C is the regularization hyperparameter and controls the trade-off between minimizing errors and penalizing overfitting. If C = 0, the algorithm disregards individual deviations and constructs the simplest hyperplane for which every observation is still within the acceptable margin of ε. For sufficiently large C, the algorithm will construct the most complex hyperplane that predicts the outcome for the training data with zero error (Bolhuis and Rayner 2020).

Chart [x]: A simple SVM to nowcast GDP



Note: The orange hyperplane is the prediction, while blue dots are actual observations. The model contains 40 regressors, while to visualize, we only use term premium and economic sentiment.

[1] Bolhuis and Rayner (2020) define support vector machine (SVM) as an algorithm that constructs hyperplanes to partition predictor combinations and make a point forecast for each of the sections.

In practice, the constraint optimization problem can be solved using Lagrange dual formulation.[2] Introducing nonnegative Lagrangian multipliers α_t and α_t^* for each observation X_t, the dual Lagrangian is as follows:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{T}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i'x_j + \varepsilon\sum_{i=1}^{T}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{T} y_i(\alpha_i^* - \alpha_i)$$

$$s.t. \quad \sum_{t=1}^{T}(\alpha_t - \alpha_t^*) = 0$$

$$\forall t: 0 \leq \alpha_t \leq C$$

$$\forall t: 0 \leq \alpha_t^* \leq C$$

Solving the problem, we obtain

$$\beta = \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)x_t$$

and consequently for each new observation x, the prediction will be

$$f(x) = \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)(x_t'x) + b.$$

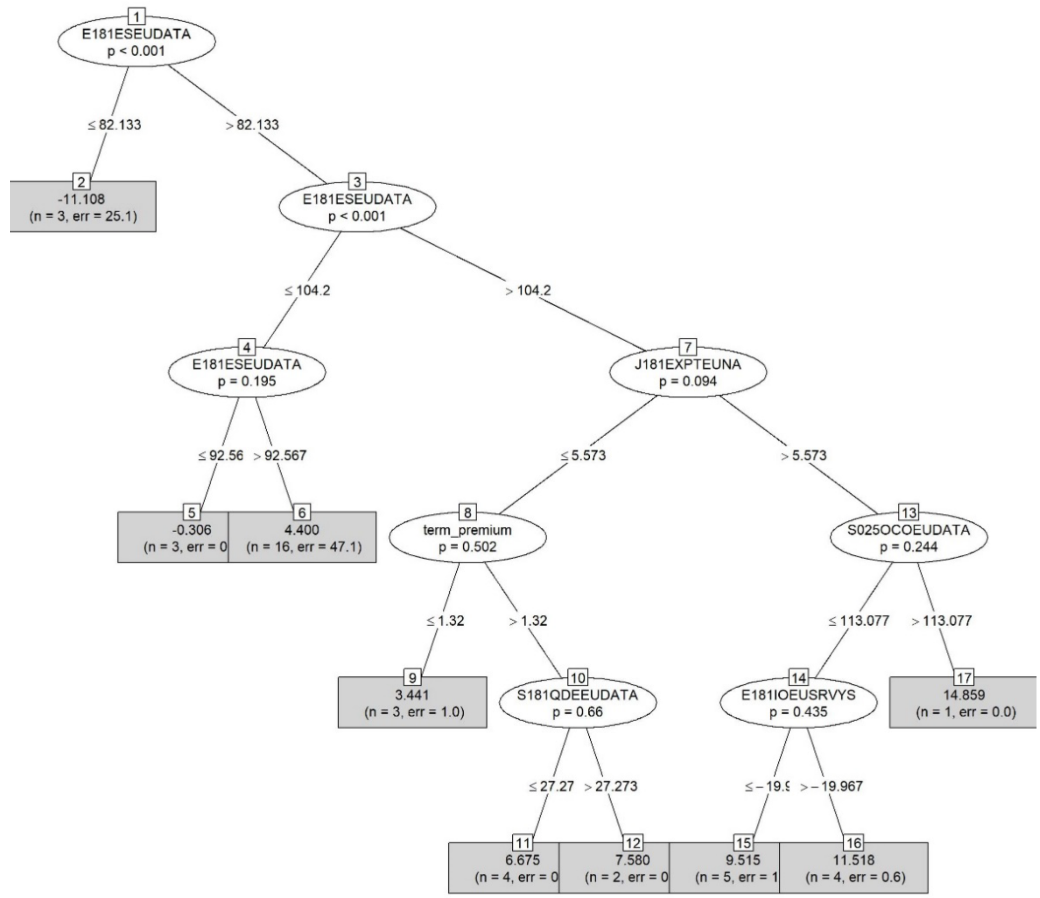
Replacing the dot product $x_j'x_k$ by other kernel functions, SVM can be applied to non-linear problems.

3. Random Forest

A decision tree is a non-parametric approach that in each step iteratively splits a sample into two groups chosen by the algorithm to yield the largest reduction in the forecast error of the variable of interest (Bolhuis and Rayner (2020)), the so-called recursive partitioning. *Regression tree* is nonparametric regression method that allows for prediction of continuous variables (Chart [x]). *Random forest* (RF) is an ensemble method based on a large number of individual decision (regression) trees created from different samples.

[2] More details can be found in Drucker and others (1997).

Chart [x]: A simple regression tree to nowcast GDP



Note: each terminal node represents a region R_m.

A formal expression of a regression tree is

$$\hat{f}(x) = \sum_{m=1}^{M} \hat{c}_m I(x \in R_m)$$

where $I(\cdot)$ is the indicator function and $\hat{c}_m = avg(y_i|x_i \in R_m)$, and R_m is regions or the groups after iteratively partitioned by the tree. The estimation then is seeking the optimized R_m and c_m to minimize the squared error:

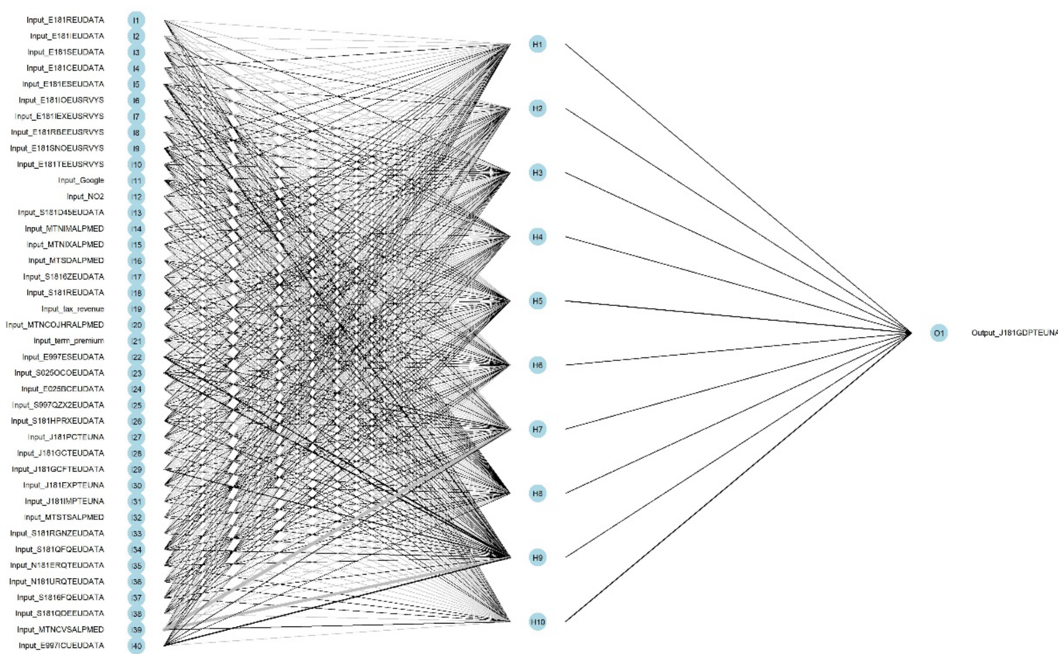
$$\min_{\{R_m, c_m\}_{m=1}^{M}} \sum \left[y_t - \hat{f}(x)\right]^2$$

A single regression tree tends to overfit data but also suffers from path dependence and model instability due to its reliance on local rather than global optimization. These drawbacks have been addressed by RF. As an ensemble method, RF uses the bootstrap aggregation (‘bagging’) to create a forest of individual trees, each of which is estimated by a randomly chosen subsample of the observations as well as the predictors. For regression tasks, the mean or average prediction of the individual trees is returned.

4. Neural Network

Neural network (NN) is a multi-layer non-linear method to map a series of inputs to a target output. Each layer contains multiple nodes called artificial neurons. Each node receives inputs either from the input data matrix or from nodes in previous layers and passes its output either to nodes in the next layer or as the final output of the model. At each node, a weighted sum of the inputs is transformed by a non-linear function $f(\cdot)$ to generate the output. Typical functions used in neural networks include the rectified linear unit (ReLU) $f(z) = \max(0, z)$, as well as hyperbolic tangent $f(z) = (e^{z} - e^{-z})/(e^{z} + e^{-z})$ or logistic function $f(z) = 1/(1 + e^{-z})$.
Given the structure of a NN, there is no closed form solution that could be estimated. Instead, NN is trained using stochastic gradient descent method, which (i) generates random weights, (ii) calculates the loss function between the target variable and the output predicted based on these random weights, and (iii) minimizes the loss function by adjusting weights with calculated gradients (LeCun Bengio and Hinton 2015).

Chart [x]: A simple neural network to nowcast GDP



Note: Each node represents a function $f(\cdot)$ transforming inputs to outputs. Each link represents the weight, while black line means positive weight, gray line means negative weight.

In a standard setting, NN is feedforward and fully connected, meaning that output from a specific node of each layer will flow unidirectionally to all other the nodes in the next layer. Given the high flexibility in designing and/or adding layers, several popular variations of NN are also introduced in our toolkit:

- *The long short-term memory* (LSTM) is a subtype of the larger class recursive neural networks (RNN), which adds feedback connections in addition to the feedforward connections in classical NN. Such modification makes LSTM more suitable for time series prediction.

- *Convolutional neural network* (CNN) introduces the so-called convolutional and pooling layers to addresses computational issues caused by fully connected layers. Compared to networks with full connectivity between adjacent layers, CNN is much easier and significantly faster to train (LeCun Bengio and Hinton 2015).

5. Indicators of predictive accuracy: Models by Country

Austria: Performance Indicators 2017Q1-2021Q1			
	Forecast Errors1/		Mean Directional
	MAE	RMSE	Accuracy2/
Benchmark			
AR1	1.93	3.75	0.50
Machine Learning Models			
ElNet_SEm	0.84	1.29	0.75
ElNet_SE1	0.86	1.27	0.69
Lasso_SEm	0.85	1.33	0.75
Lasso_SE1	0.87	1.30	0.69
Ridge_SEm	0.82	1.26	0.75
Ridge_SE1	0.82	1.23	0.69
RFor	1.74	3.03	0.75
SVMachine	1.04	1.65	0.69
LinReg	0.99	1.38	0.63
GPReg	1.48	2.76	0.56
NN	1.33	1.82	0.75
NNdeep	1.49	2.40	0.56
NNlstm	1.48	3.01	0.75
NNbilstm	1.61	2.78	0.56
NNconv	1.41	2.11	0.63
Dynamic Factor Model			
DFM	1.19	1.88	0.63
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

Hungary: Performance Indicators 2016Q1-2021Q1			
	Forecast Errors		Mean Directional
	MAE	RMSE	Accuracy2/
Benchmark			
AR1	2.70	3.80	0.67
Machine Learning Models			
ElNet_SEm	1.98	1.22	0.71
ElNet_SE1	2.06	1.27	0.67
Lasso_SEm	1.96	1.21	0.81
Lasso_SE1	2.09	1.30	0.67
Ridge_SEm	2.03	1.20	0.67
Ridge_SE1	2.03	1.24	0.67
RFor	2.72	3.16	0.86
SVMachine	2.14	1.39	0.67
LinReg	2.22	1.56	0.76
GPReg	1.84	2.71	0.76
NN	2.03	2.07	0.57
NNdeep	2.01	2.02	0.71
NNlstm	2.21	2.57	0.52
NNbilstm	2.07	2.13	0.57
NNconv	2.63	2.22	0.62
Dynamic Factor Model			
DFM	2.16	2.11	0.76
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

Ireland: Performance Indicators 2015Q2-2021Q1			
	Forecast Errors		Mean Directional
	MAE	RMSE	Accuracy2/
Benchmark			
AR1	5.48	7.73	0.61
Machine Learning Models			
ElNet_SEm	10.93	20.34	0.48
ElNet_SE1	10.93	20.34	0.48
Lasso_SEm	11.11	20.13	0.57
Lasso_SE1	11.18	20.11	0.57
Ridge_SEm	10.84	20.44	0.52
Ridge_SE1	10.84	20.44	0.52
RFor	5.44	7.86	0.65
SVMachine	8.90	15.87	0.57
LinReg	9.55	14.44	0.57
GPReg	5.46	6.62	0.57
NN	6.14	8.09	0.70
NNdeep	6.24	7.57	0.61
NNlstm	6.56	8.39	0.52
NNbilstm	6.23	7.59	0.48
NNconv	7.32	10.46	0.48
Dynamic Factor Model			
DFM	5.44	7.13	0.61
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

Malta: Performance Indicators 2012Q1-2021Q1			
	Forecast Errors		Mean Directional
	MAE	RMSE	Accuracy2/
Benchmark			
AR1	5.15	3.57	0.64
Machine Learning Models			
ElNet_SEm	3.17	3.67	0.61
ElNet_SE1	5.56	3.62	0.58
Lasso_SEm	5.87	3.61	0.61
Lasso_SE1	5.82	3.59	0.58
Ridge_SEm	5.88	3.74	0.64
Ridge_SE1	5.47	3.63	0.61
RFor	3.94	4.35	0.67
SVMachine	3.83	4.14	0.67
LinReg	3.57	5.57	0.58
GPReg	5.10	4.19	0.72
NN	5.51	3.94	0.56
NNdeep	5.38	3.32	0.58
NNlstm	4.15	5.06	0.61
NNbilstm	4.93	4.55	0.58
NNconv	4.87	4.50	0.78
Dynamic Factor Model			
DFM	5.20	2.77	0.67
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

Poland: Performance Indicators 2015Q2-2021Q1			
	Forecast Errors		Mean Directional
	MEA	RMSE	Accuracy2/
Benchmark			
AR1	1.00	2.06	0.63
Machine Learning Models			
ElNet_SEm	0.94	1.64	0.75
ElNet_SE1	0.92	1.61	0.75
Lasso_SEm	0.99	1.71	0.72
Lasso_SE1	0.94	1.67	0.72
Ridge_SEm	0.90	1.45	0.72
Ridge_SE1	0.89	1.45	0.72
RFor	1.19	1.80	0.69
SVMachine	0.90	1.60	0.78
LinReg	1.31	2.20	0.75
GPReg	1.10	1.53	0.66
NN	1.18	1.98	0.66
NNdeep	1.22	2.25	0.72
NNlstm	1.31	2.22	0.72
NNbilstm	1.56	2.45	0.66
NNconv	1.23	1.96	0.84
Dynamic Factor Model			
DFM	1.13	1.71	0.72
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

Portugal: Performance Indicators 2015Q3-2021Q1			
	Forecast Errors		Mean Directional
	MAE	RMSE	Accuracy2/
Benchmark			
AR1	1.71	3.95	0.50
Machine Learning Models			
ElNet_SEm	0.64	1.18	0.77
ElNet_SE1	0.64	1.18	0.77
Lasso_SEm	0.65	1.20	0.77
Lasso_SE1	0.62	1.17	0.77
Ridge_SEm	0.67	1.25	0.77
Ridge_SE1	0.67	1.25	0.77
RFor	1.56	3.02	0.68
SVMachine	0.77	1.62	0.73
LinReg	0.79	1.53	0.73
GPReg	1.31	3.62	0.77
NN	1.66	3.05	0.73
NNdeep	1.32	2.56	0.64
NNlstm	1.12	2.69	0.68
NNbilstm	1.28	2.87	0.64
NNconv	0.96	1.71	0.73
Dynamic Factor Model			
DFM	1.48	2.80	0.55
1/ Lower values indicate better performance.			
2/ Higher values indicate better performance.			

References

Antolin-Diaz J., Drechsel T., and Petrella I., 2021. “Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data”, CEPR Discussion Papers 15926, C.E.P.R. Discussion Papers.

Amat C., Michalski T., and Stoltz G., 2018. "Fundamentals and exchange rate forecastability with simple machine learning methods," Journal of International Money and Finance, Elsevier, vol. 88(C), pages 1-24.

Athey S., (2019), “The Impact of Machine Learning on Economics”, in The Economics of Artificial Intelligence: An Agenda, University of Chicago Press.

Bańbura M., Giannone, D., Modugno, M. and Reichlin, L., 2013. Now-casting and the real-time data flow. In Handbook of economic forecasting (Vol. 2, pp. 195-237). Elsevier.

Bańbura M., Modugno M. (2010) Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. European Central Bank, Working Paper 1189

Basuchoudhary A, Bang J. T., and Sen T., 2014. “Machine-learning Techniques in Economics - New Tools for Predicting Economic Growth”, ISSN 2191-5512 (electronic), SpringerBriefs in Economics.

Belloni A., Chernozhukov V., and Hansen C., 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects," Journal of Economic Perspectives, American Economic Association, vol. 28(2), pages 29-50, Spring.

Bok, Brandyn and Caratelli, Daniele and Giannone, Domenico and Sbordone, Argia M. and Tambalotti, Andrea, Macroeconomic Nowcasting and Forecasting with Big Data (2017-11-01). FRB of NY Staff Report No. 830

Bolhuis, M.A., Rayner, B., 2020. “Deus ex Machina? A Framework for Macro Forecasting with Machine Learning.” IMF Working Paper No. 20/45, Washington, DC.

Camacho M., Perez-Quiros G., 2010."Introducing the euro-sting: Short-term indicator of euro area growth," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 25(4), pages 663-694.

Carvalho, D.V., Pereira, E.M. and Cardoso, J.S., 2019. Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), p.832.

Chapman J. and Desai A., 2021, “Using Payments Data to Nowcast Macroeconomic Variables During the Onset of COVID-19”, Staff Working Papers 21-2, Bank of Canada.

Choi H. and Varian H., 2009a, “Predicting the Present with Google Trends”, The Economic Record, The Economic Society of Australia, vol. 88(s1), pages 2-9, June.

Choi H. and Varian H., 2009b, “Predicting initial claims for unemployment insurance using Google Trends.” Technical report, Google, 2009b.

Corona F., Gonzalez-Farias G., and Jesus Lopez-Perez, 2021. “A nowcasting approach to generate timely estimates of Mexican economic activity: An application to the period of COVID-19”, Papers 2101.10383, arXiv.org.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A. and Vapnik, V., 1997. Support vector regression machines. Advances in neural information processing systems, 9, pp.155-161.

Fagan, G. and Fell J., 1991. “Tecniques for short-term economic forecasting”, Journal of the Statistical and Social Inquiry Society of Ireland, Vol XXVI, Part III.

Gareth J., Witten D., Hastie T., Tibshirani R., 2015. An Introduction to Statistical Learning : with Applications in R. New York :Springer, 2015.

Hastie, T., Tibshirani, R., and Friedman, J. H., 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer.

Hastie, T., Tibshirani, R. and Wainwright, M., 2015. “Statistical Learning with Sparsity: The Lasso and Generalizations.” CRC Press.

Hopp, D., 2021. Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM). arXiv preprint arXiv:2106.08901.

Hutter, Christian, 2020. A new indicator for nowcasting employment subject to social security contributions in Germany, Journal for Labour Market Research, Vol. 54, Issue. 1.

Kohlscheen, E., 2021.”What Does Machine Learning Say About the Drivers of Inflation?” , BIS Working Paper 980.

Knotek, E. S. II, and Saeed Zaman, 2014, Nowcasting U.S. Headline and Core Inflation, Federal Reserve Bank of Cleveland, Working Paper no. 14-03.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

Lewis, D., K. Mertens, J.H. Stock, and M. Trivedi. 2020, Measuring Real Activity Using a Weekly Economic Index, Federal Reserve Bank of New York Staff Reports, no. 920, April, revised September.

Lewis, D., K. Mertens, J.H. Stock, and M. Trivedi. 2020, High Frequency Data and a Weekly Economic Index during the Pandemic, Federal Reserve Bank of New York Staff Reports, no. 954, December.

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S., 2021. Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), p.18.

Marlin, B., 2008. Missing data problems in machine learning (Doctoral dissertation).

Modugno, Michele, 2011, Nowcasting Inflation Using High Frequency Data, ECB Working Paper No. 1324

Muchisha, N. D. ., Tamara, N. ., Andriansyah, A., and Soleh, A. M., 2021. “Nowcasting Indonesia’s GDP Growth Using Machine Learning Algorithms”. Indonesian Journal of Statistics and Its Applications, 5(2), 355–368.

Reichlin, L., Giannone, D., Banbura, M., "Nowcasting", in Michael P. Clements and David F. Hendry, editors, Oxford Handbook on Economic Forecasting, 2011

Richardson, A., van Florenstein Mulder, T. and Vehbi, T., 2021. “Nowcasting GDP Using Machine-Learning Algorithms: A Real-time Assessment.” International Journal of Forecasting, 37(2), pp.941-948.

Sampi B.,E. and Jooste,C., 2020. "Nowcasting Economic Activity in Times of COVID-19 : An Approximation from the Google Community Mobility Report," Policy Research Working Paper Series 9247, The World Bank.

Storm H., Baylis K., and Heckelei T., 2020."Machine learning in agricultural and applied economics," European Review of Agricultural Economics, Foundation for the European Review of Agricultural Economics, vol. 47(3), pages 849-892.

Tiffin, A.J., 2016. “Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon.” IMF Working Paper No. 16/56, Washington, DC.

Tuhkuri J., 2016, "ETLAnow: A Model for Forecasting with Big Data – Forecasting Unemployment with Google Searches in Europe," ETLA Reports 54, The Research Institute of the Finnish Economy.

Varian H., 2014, “Big Data: New Tricks for Econometrics”, Journal of Economic Perspectives—Volume 28, Number 2—Spring 2014—Pages 3–28

Woloszko, N., 2020. “Tracking Activity in Real Time with Google Trends (No. 1634).” OECD Publishing.

PUBLICATIONS

Nowcasting GDP

A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies

Working Paper No. **WP/22/52**