IDB
Inter-American Development Bank

Nowcasting to predict economic activity in real time: The cases of Belize and El Salvador

Juan Barrios
Inter-American Development Bank

Lucía Martín
Inter-American Development Bank

Julia Escobar
Inter-American Development Bank

Werner Peña
University of Kent

Janelle Leslie
Inter-American Development Bank

Belize and El Salvador

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Nowcasting to predict economic activity in real time: the cases of Belize and El Salvador / Juan Barrios, Julia Escobar, Janelle Leslie, Lucía Martín, Werner Peña.
p. cm. — (IDB Monograph ; 970)
Includes bibliographic references.

1. Economic forecasting-Belize. 2. Economic forecasting-El Salvador. 3. Machine learning-Belize. 4. Machine learning-El Salvador. I. Barrios, Juan José. II. Escobar, Julia. III. Leslie, Janelle. IV. Martín Rivero, Lucía. V. Peña, Werner. VI. Inter-American Development Bank. Country Department Central America, Haiti, Mexico, Panama and the Dominican Republic. VII. Series.

IDB-MG-970

JEL Codes:
E01 Measurement and Data on National Income and Product Accounts and Wealth • Environmental Accounts, E27 Forecasting and Simulation: Models and Applications, C53 Forecasting and Prediction Methods • Simulation Methods

Key Words:
Belize, El Salvador, Nowcasting, Forecast of Economic Activity, Machine Learning, Quarterly GDP, national statistics

http://www,iadb.org

Copyright © 2021 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.
Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.
Note that link provided above includes additional terms and conditions of the license.
The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

Belize and El Salvador

Nowcasting to predict economic activity in real time: The cases of Belize and El Salvador

Juan Barrios
Inter-American Development Bank

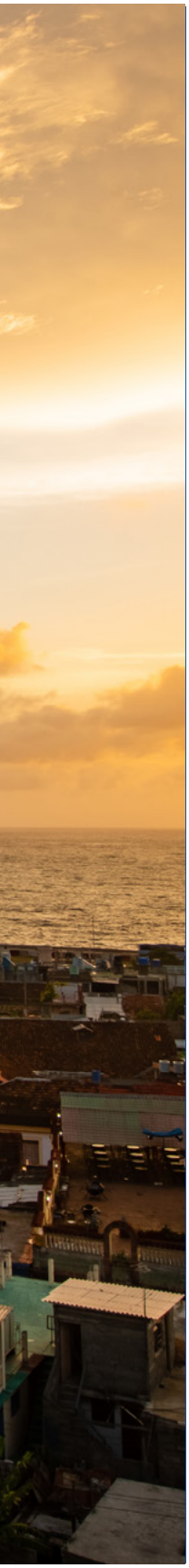
Lucía Martín
Inter-American Development Bank

Julia Escobar
Inter-American Development Bank

Werner Peña
University of Kent

Janelle Leslie
Inter-American Development Bank

Belize and
El Salvador



Contents

We are grateful to Emory Ford and Rumile Arana from the Central Bank of Belize (CBB) for giving us access to sectoral production statistics. In addition, we thank staff from the Research Department of the CBB and staff from the Statistical Institute of Belize (SIB), as well as Maria Deza (IADB) for useful comments and suggestions during preliminary presentations of this exercise. All possible errors remain with the authors.

Belize and
El Salvador

Executive Summary

Short-term forecasting capacity and real-time decision making under a high degree of uncertainty have gained unprecedented significance amid the COVID-19 pandemic. Over the past year, governments have been challenged to react quickly and effectively in order to tackle the effects of the pandemic in turbulent conditions. Those circumstances have highlighted the need for instruments that make it possible to nimbly anticipate which direction an economy might take.

A prominent feature of this uncertainty is that many key statistics are published with a substantial lag, are subsequently revised, and are available at different intervals. As a result, policymakers, economists, and business professionals—who are tasked with tracking potential economic changes—have to examine a large number of relevant variables in order to devise accurate forecasts.

In those circumstances, nowcasting techniques offer an effective tool to fill the gap arising from the lag in the release of macroeconomic indicators, by exploiting the availability of other indicators that are disseminated in shorter intervals. These techniques are based on using variables that are correlated to quarterly GDP, and that are published weekly, monthly or quarterly, so as to obtain real-time signals of movements in output and produce an accurate forecast of its value in the quarter of interest.

This paper presents machine learning models fitted to predict quarterly GDP for Belize and El Salvador. The initiative is part of an effort made by the Inter-American Development Bank (IDB) to develop timely economic monitoring tools since the onset of the pandemic. The results show that machine learning techniques can produce accurate quarterly GDP forecasts for two contrasting economies in an economic context marked by a high degree of volatility at both the national and international levels.

The model with the greatest predictive efficiency for Belize foresaw an annual GDP decline of 22.9 percent in the second quarter of 2020, as against the observed figure of 23.9 percent; a decline of 15.0 percent in the third quarter, compared to the observed figure of 12.8 percent; and a 14.2 percent contraction in the final quarter of 2020, versus the actual figure of 13.1 percent. In the case of El Salvador, the best model foresaw a contraction of 17.8 percent as against an observed rate of -19.4 percent in the second quarter of 2020; a decline of 8.0 percent compared to the observed 9.9 percent in the third quarter; and a decline of 2.3 percent in the last quarter of the year versus the observed figure of -2.3 percent. For the first quarter of 2021, the models predict a year-on-year decline of 6.8 percent for Belize (compared to the observed rate of -8.4 percent) and year-on-year growth of 1.2 percent for El Salvador (against an observed rate of 3.0 percent).

Because the calibration of nowcasting exercises is a dynamic process that is refined over time, at the Inter-American Development Bank we trust that this document will help support the ongoing work of the governments and statistical agencies of Belize and El Salvador in securing better economic forecasts to guide policy decisions.

Introduction

In January 2020, the International Monetary Fund (IMF, 2020) projected that the world economy would grow by 3.3 percent in that year, given the positive performance of the markets and the expectation of improved trade relations between China and the United States. A modest recovery was projected for Latin America and the Caribbean (LAC), with expected growth of 1.6 percent compared to 0.1 percent in 2019. Just three months later, the World Health Organization (WHO) declared that the COVID-19 outbreak had escalated to pandemic status. With that, the world underwent one of the most radical changes of the past century, a shift that caught governments off guard and portended unknown social and economic implications.

More than a year into the pandemic, the economic contraction has been severe, as have the deleterious effects on public finances. LAC governments have faced a historic health crisis, and in response, they have imposed strict and prolonged containment measures in view of their fragile health systems. Regional GDP is estimated to have fallen by 7 percent in 2020, almost four times more than during the 2009 financial crisis. At the same time, the collapse of tax revenues and a sharp increase in spending[1] to cope with the pandemic have led to widening fiscal deficits, and increased debt costs.

The uncertainty spurred by the pandemic, and the challenge of reacting quickly to cope with its effects through support programs in a turbulent environment, have underscored the need for effective monitoring tools. A year after the outbreak of the pandemic, there is still uncertainty about the dangers to the macroeconomic outlook in the medium and long terms. Containment measures have been relaxed, for instance, but the risk of further outbreaks will dominate policymaking until vaccination becomes ubiquitous. This will determine the pace at which economic activity resumes, and with it the return of sustainability to public finances.

In this context, the Inter-American Development Bank (IDB) and the Country Department Central America (CID), Haiti, Mexico, Panama, and the Dominican Republic have been working since the start of the pandemic to develop tools that offer a clearer view of the economic outlook. To that end, the IDB has calibrated machine learning models to produce nowcasting exercises of quarterly GDP in two pilot countries. One of the main advantages of these exercises is that they make it possible to predict the behavior of economic variables such as quarterly GDP given the availability of high-frequency data, thereby filling the gap between the end of a period (usually a quarter) and the publication of the variable in question by the statistical authorities.

The aim of this paper is to present the machine learning models used in nowcasting exercises, as well as their application and results in two Central American countries, Belize and El Salvador. The IDB thereby seeks to help facilitate decision-making by providing their governments and key stakeholders with timely and valuable input on the direction of the economy in the two countries. In Belize, the Statistical Institute of Belize (SIB) publishes quarterly GDP about two months after the end of each quarter, while the Central Reserve Bank of El Salvador publishes it 90 days after the quarter ends. In both cases, nowcasting quarterly GDP can make up for the lack of information in that 60–90 day gap.

Apart from producing substantially accurate quarterly GDP forecasts for both countries, the results presented herein highlight the ability of machine learning models to produce fore-

1 The main measures included, among others, supplementary budget allocations, cash transfers to households and businesses, relaxation of fiscal rules, cutting or deferring tax payments, and direct payment of part of the corporate payroll.

casts in an economic context marked by a high degree of volatility. An important advantage of these models, moreover, is that they are better suited to learning how to recognize downturns in GDP that have not been observed in the past. In doing so, they can produce more accurate results than are available using standard econometric techniques.

It is important to keep in mind, however, that these exercises depend on the quality of the preliminary data used for the forecast. In other words, the immediate forecast is based on the assumption that the preliminary figures do not change drastically after their revision. The results also confirm that machine learning models are a sound option to produce accurate exercises in immediate forecasting for small economies that still face some challenges in data collection and availability.

This paper is organized into eight sections. Following this introduction, the second section explains why Belize and El Salvador were chosen as the two pilot countries, as well as the statistical capacity available in each of them. The third section defines the machine learning techniques used in the nowcasting exercises. The fourth section examines in detail each of the machine learning methods used in the study. The fifth section describes the data used in each country and the associated peculiarities. The sixth section presents the results of the prediction methods and compares their relative predictive efficiency. The seventh section details the collaboration with the authorities and institutional strengthening in the pilot countries. Finally, the eighth section presents some conclusions and lessons learned from using machine learning tools to predict quarterly GDP.

nowcasting techniques offer an effective tool to fill the gap arising from the lag in the publication of macroeconomic indicators

Belize and
El Salvador

Selection of Belize and El Salvador

Belize and El Salvador were chosen as the two pilot countries for this exercise from among the CID countries in order to test the effectiveness of these methods in two economies that, though similar in certain respects, also feature significant structural differences. While the measures implemented in practically all countries to address the health and economic crises have implications that require frequent and timely measurement, Belize and El Salvador were chosen as a means of testing the models' effectiveness in producing reliable results in all countries, irrespective of their particular characteristics.

The following subsections briefly describe the socioeconomic contexts in Belize and El Salvador, so that readers can compare the two countries' similarities and differences. Essentially, although both countries have small, open economies and are located in Central America, there is a substantial difference in the size of their economies, and Belize has historical and commercial ties with the Caribbean that El Salvador does not share. The most significant differences include, among others, the economic sectors that form the basis of their productive apparatus, their trade partners, their fiscal conditions, the financing sources available to them, the effects of the pandemic, and the challenges they each face. Even more importantly, there are substantial differences in their statistical capacities. Despite this, and as is evident in the fifth section of this paper, the results of the quarterly GDP forecasting confirm that the methodology used can yield reliable results in economies that are structurally different.

Belize: socioeconomic context

Belize is a small, open economy in Central America that relies heavily on the production and export of commodities and tourism. In 2019, Belize's GDP was US$1.9 billion, the smallest among IDB member countries. The population was a little above 417,000, the third smallest among IDB members.[2] Tourism is the leading economic sector, accounting for about 37.2 percent of GDP in 2019.[3] Tourism is responsible for 39.3 percent of total employment and 48 percent of total exports of goods and services. Before the pandemic, about 70 percent of visitors to Belize came from the United States. The second most important sector is agriculture, which accounts for about 10 percent of GDP, employs more than 15 percent of the active labor force, and produces more than 80 percent of goods exports. The United Kingdom is the main destination for goods exports, taking 34 percent of the total in 2020, followed by the United States (22 percent) and CARICOM countries (17 percent).

Because of its economic dependence on tourism, Belize has been among the countries most affected by the pandemic worldwide. In 2020 the economy contracted by 14.1 percent, the largest drop in economic activity in the country's history, caused by a 70 percent year-on-year decline in tourism activity and restrictions imposed on internal mobility. This unprecedented blow added pressure to an economy whose growth has been low in recent decades, bringing the size of the economy, in constant terms, to 2011 levels. Per capita income fell from US$4,699 in 2019 to US$3,944 in 2020, and unemployment is estimated at 13.7 percent of the employed population. During

2 After Barbados and The Bahamas.

3 According to the World Travel and Tourism Council (WTTC). These percentages cover the direct and indirect effects of tourism. Figures correspond to the year 2019.

the pandemic the country was struck by Hurricanes Eta and Iota, which caused damages equivalent to about 0.4 percent of GDP and aggravated the country's economic and social challenges.

As a result of the impact of the pandemic and the hurricanes, in 2020 Belize registered record fiscal shortfalls and an increase in its debt. The primary and overall deficits are estimated at 8.3 percent and 10.8 percent of GDP in 2020, respectively. Belize does not have a fiscal rule, but repayment of its external bond debt is linked to meeting fiscal targets. Its non-compliance since 2019 has triggered the quarterly, rather than semiannual, payment of "Superbond" coupons. The level of indebtedness rose abruptly from 97.5 percent of GDP in 2019 to 125.8 percent in 2020. The country's delicate fiscal conditions prompted a downgrade by S&P (to CC) and by Moody's (to Caa3), as well as a leap up in the Emerging Market Bond Index (EMBI) from 1,042 basis points in April to 1,600 in July 2020 (it is currently at 1,406).

El Salvador: socioeconomic context

El Salvador's economy (US$27 billion in 2019) is 14 times bigger than Belize's, though it is still considered a small economy compared to other Central American countries such as Costa Rica and Panama, or the Dominican Republic.[4] The Salvadoran economy is highly dependent on the performance of the United States and on remittances, is only modestly diversified, and is dominated by the tertiary sector (70 percent of GDP),[5] in which two of every three workers are informal (DIGESTYC, 2020).

The country has experienced historically weak growth, a circumstance that has kept it in the lower-middle income category for 30 years[6] and has limited the economic opportunities available to the population. It has also endured a constant increase in social and fiscal vulnerability,[7] which has triggered a high level of emigration of thousands of Salvadoran citizens, mainly to the United States. In the fiscal arena, despite the gradual fiscal consolidation of the past decade, the level of public indebtedness is high (Ministerio de Hacienda, 2021)[8] and the cost of financing is rising. This debt situation, coupled to low economic growth, constrains the fiscal maneuvering room to deal with external shocks and natural disasters.

As a result of the pandemic containment measures and the external shock of lower growth in the United States, the Salvadoran economy contracted by 7.9 percent in 2020, one of the sharpest declines in Central America. Goods exports suffered an annual fall of 15 percent, similar to the export decline during the 2009 financial crisis. Family remittances provided an important cushion. They fell sharply by about 40 percent in the first months of the pandemic but, as the United States granted fiscal packages in the form of economic relief to households, remittances began to recover and achieved growth of 4.7 percent by year's end. Despite the recovery of remittances, the contraction of production and formal employment contributed to a rise in poverty (DIGESTYC, 2021) and inequality (IDB, 2021).

The economic downturn and the budgetary measures adopted to meet demands in the areas of health and economic relief have affected fiscal conditions. In 2020 the fiscal deficit tripled to 10 percent of GDP, while debt increased by 18 percentage points, reaching 87.9 percent of GDP by the end of the year (Ministerio de Hacienda, 2021).

Statistical capacity of Belize and El Salvador

A crucially important aspect of nowcasting exercises is the quantity and quality of information available to calibrate the models. In the case of quarterly GDP, this exercise requires high-frequency information (monthly or weekly data, for example) on variables related to quarterly movements in output.

In Belize, data availability poses a major challenge. The Statistical Institute of Belize (SIB) publishes quarterly GDP data about two months after the end of each quarter, and regularly publishes data on foreign trade and prices. The Central Bank of Belize (CBB) publishes regular data on monetary and financial aggregates, foreign trade, and sectoral output. The Belize Tourism Board (BTB) records monthly up-

4 El Salvador accounts for 10.6 percent of the population and 7.5 percent of the GDP of the region of Central America, Panama, and the Dominican Republic.

5 Central Reserve Bank of El Salvador (BCR).

6 The World Bank classifies the world's economies into four per-capita income groups: high, upper-middle, lower-middle and low. In 1987, El Salvador was together with 18 LAC economies in the lower-middle income group. By 2019, El Salvador was one of only four economies that had not moved to a higher income level.

7 Before the Covid-19 crisis, two of every three Salvadorans were poor or were at risk of falling into poverty in the face of negative changes in growth and/or remittances.

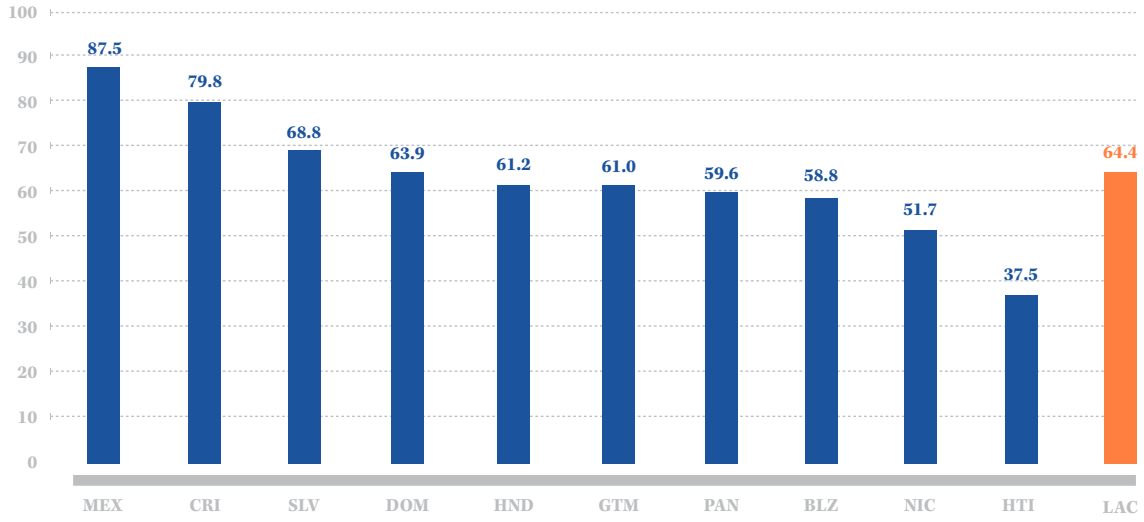
8 In 2020, debt was equivalent to 85.8 percent of GDP, and for 2021 it is estimated to rise to 88.6 percent of GDP.

dates on arrivals and departures into and from the country. Despite these statistical capacities, in areas such as sectoral output or monthly fiscal statistics (such as tax revenues), the monthly information is published at intervals that limit its availability for real-time nowcasting. Similarly, there is no public access to monthly data on the number of people working in the formal sector, nor monthly information on energy production and consumption; in nowcasting exercises, these variables generally have a strong association with quarterly GDP. [9]

In El Salvador, every month the Central Reserve Bank publishes a variety of output indices, foreign trade statistics, and fiscal statistics. The Salvadoran Social Security Institute (ISSS) publishes monthly data on the number of active contributors (a proxy for formal employment) with a lag of one or one and a half months, allowing this information to be included in the quarterly output forecast.

The statistical capacities outlined above are reflected in the World Bank's statistical performance indicator for several Central American countries, in which El Salvador ranks third and Belize ranks eight. In this regard, nowcasting exercises in both countries must take account of the constraints and challenges attendant on their statistical capacities.

STATISTICAL PERFORMANCE INDICATOR



Source: Overall level of statistical capacity. World Bank (2021).

9 Note that, in Belize, the Statistical Institute is currently working on estimating a new GDP series, which is expected to be published in the coming months. The models presented here can easily be recalibrated to produce nowcasting exercises with the forthcoming series.


IDA
MACIAS

Nowcasting for timely information

The term nowcasting is a contraction of "now" and "forecasting," which are combined to make it explicit that the aim of this kind of exercise is to forecast an economic variable in real time when the variable has not yet been estimated and/or published (Tiffin, 2016). This is precisely why nowcasting exercises have attracted so much attention recently, especially from policymakers, statistical institutes, and central banks. Macroeconomic variables such as GDP are published with a lag of two to three months in most LAC countries, which somewhat restricts the information available to policymakers and all stakeholders for short-term decision-making. Efforts to assess the state of the economy in real time therefore resort to data that were published in the past and that do not reflect current conditions—a suboptimal undertaking in circumstances of high volatility.

Nowcasting exercises make it possible to fill the gap arising from the lag in the publication of macroeconomic indicators, by exploiting the availability of other indicators that are released more often. These models are based on using variables that are semi-correlated or highly correlated to GDP, and that are published weekly, monthly or quarterly, so as to secure real-time signals of movements in GDP and produce an accurate forecast of the value of output in the period of interest.

With the development of machine learning tools (which can be defined as a statistical application that learns how to improve its own performance in order to produce accurate results), nowcasting exercises based on supervised algorithms have gained popularity. These supervised learning algorithms aim to capture the strength of the relationship between the target variable and a set of predictors, exploiting the characteristics of the predictors and their interaction with the target variable. Among these supervised learning techniques, the most widely used include penalized regressions and decision trees. Both techniques have been used to predict quarterly GDP.

For example, Tiffin (2016) uses penalized regressions and decision based algorithms (random forest) to nowcast quarterly GDP for Lebanon. Jung et al. (2018), Bolhuis and Rayner (2020a), and Bolhuis and Rayner (2020b) discuss the use of machine learning techniques, including penalized regressions and methods based on decision trees, to predict macroeconomic variables. These authors highlight the following advantages of using machine learning techniques to predict quarterly GDP: (i) machine learning models prioritize out-of-sample performance; (ii) these models are better suited to dealing with nonlinearities in the data; and (iii) given their flexibility, they can model complex relationships among variables.[10]

In this paper, machine learning techniques are used to nowcast quarterly GDP in Belize and El Salvador.

10 These characteristics allow machine learning models to avoid the classic problems present in traditional immediate-prediction methods, including: the presence of collinearity between predictors, differences in data dimension, relevance of predictors, and the presence of nonlinearities (Bolhuis and Rayner, 2020b).


ATACO

Machine learning techniques for nowcasting exercises

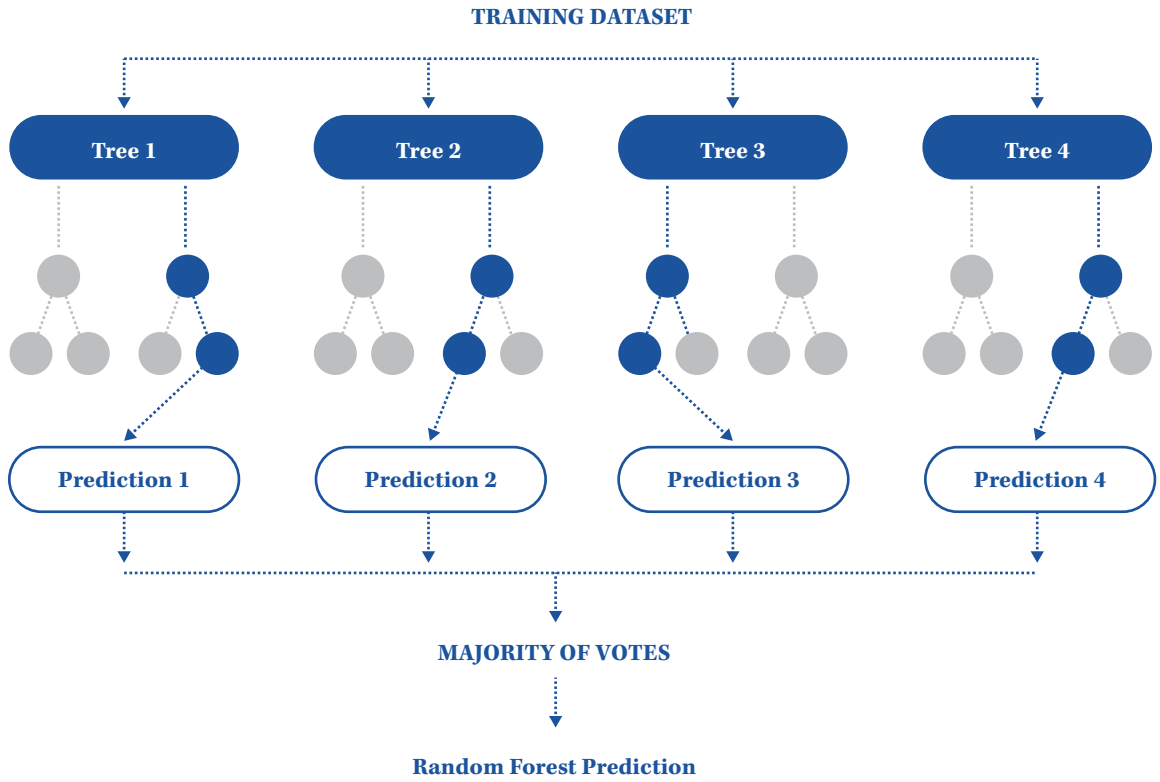
The following is a brief description of the methods used to nowcast quarterly GDP for Belize and El Salvador. The six methods used in the exercises are: random forest, gradient boosting machine (GBM), neural networks, lasso regression, ridge regression, and elastic net regression. Note that this section presents the methods in a very general way; see the relevant literature for an exhaustive discussion of them.

Random forest

This method is based on building decision trees from the variables contained in a matrix X and a random selection of attributes. In simple terms, the random forest technique involves a uniform random selection with replacement of different subsamples or N cases of the data contained in a matrix X, which serve to train and grow each tree that is part of the forest. Random forest differs from other techniques based on decision trees in that at each node there is a random selection of variables or features, which in turn are used to split each node (Breiman, 2001). In this regard, each tree produces a prediction of the target variable, in our case quarterly GDP, and as a whole the model chooses the best prediction, which is precisely the one that secures the "most votes" in the forest of decision trees (bootstrap aggregation) (see Figure 1). Compared to other methods, random forest has the advantage of accumulating the predictions of the trees that make up the forest, which—together with the selection of the trees with the lowest error rate—confers a good degree of protection against the errors that each tree can produce individually (provided the correlation among the trees is low). The random forest method thus recursively divides the data contained in X into a series of regions, $X_1, X_2, X_3 \ldots X_r$, which, together with a random selection of predictors, will produce a forecast of the outcome variable. Then, in order to secure optimal predictions, each region of X must be the one that optimizes the purity of the dataset in line with the classes K. Then, the dependent variable is obtained as the average for the regions using the following expressions (Tiffin, 2016):

$$\hat{f}(x) = \sum_{m} \hat{c}_m I(x \in X_r)\,;\ \hat{c}_m = avg(y_i | x_i \in X_r) \qquad (1)$$

Figure 1. Simplified Representation of the Random Forest Technique



Source: Adapted from Chakure (2019).

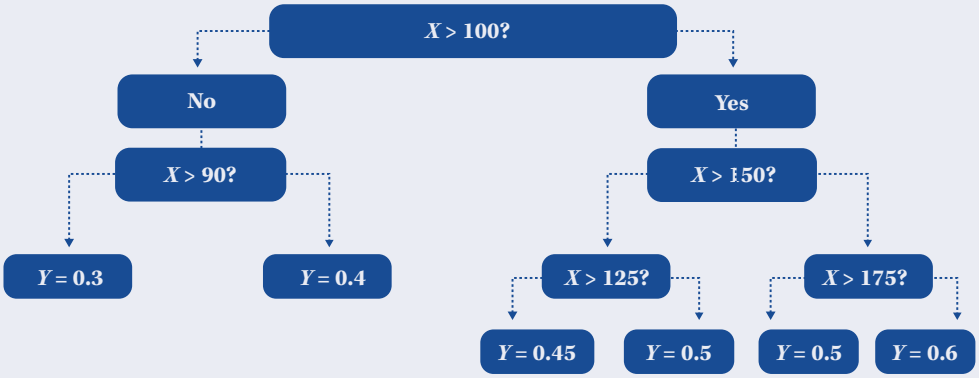
According to Breiman and Cutler (n.d.), the advantages of the random forest technique include the following: (i) it is efficient in handling large databases with a high number of variables; (ii) it provides an estimate of which variables are important in the classification; and (iii) it yields an unbiased internal estimate of the generalization error as the construction of the forest progresses. Among the main disadvantages of this technique are: interpreting the results beyond forecasts is complicated; and the training, especially the tuning of hyperparameters, is computationally demanding. To calibrate this model, the hyperparameters were fine-tuned and the cross-validation technique was used to train the model and secure the best out-of-sample performance.

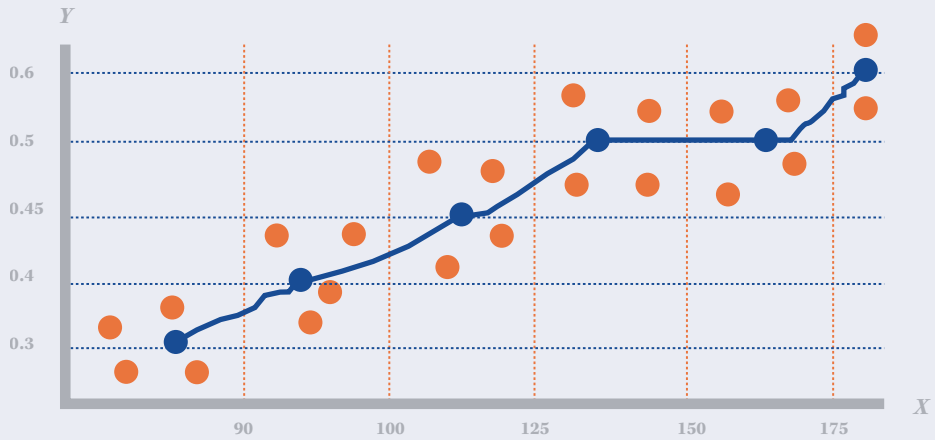
Box 1:
Regression Tree

Following Tiffin (2016), it is important to define the functioning of the regression trees that form the basis of the random forest and gradient boosting machine techniques (as will be seen later). The figures in this box help explain how they work. Suppose we have a predictor X and a response variable Y in a database with N observations. Initially, the data are in the same cluster or database. The regression tree starts with the first partition of the database on the

basis of the characteristics of the information contained in the database, and particularly on the regressor variable (or variables). Suppose that, in this first step, 100 is a point that allows us to clearly separate the information contained in the database into two groups (taking into account the minimization of the overall deviation from the mean in each of the partitions). This first partition gives rise to subsequent partitions of the information, based on rules that allow the subgroups to be separated into new branches until a minimum node size is reached, which is determined *a posteriori* (in the figures below, the prediction of the regression tree (superior panel) is presented in the inferior panel). The regression trees thus estimate the value of the response variable by obtaining the mean value of the variable for each node. With regard to the size of the tree (the number of node shares), a balance must be struck between in-sample and out-of-sample fit, because growing the trees as a whole can lead to an overfitting that gives primacy to the in-sample fit over the out-of-sample fit, yielding inaccurate forecasts. To avoid overfitting, the regression trees are optimally "pruned" to improve the out-of-sample forecast at the expense of in-sample performance.

Source: Tiffin (2016).





Box 2:
Cross-Validation

Cross-validation is a technique that allows us to assess and optimize the probable out-of-sample performance of the calibrated model. Essentially, the technique consists of dividing the database we are working with into different folds, and choosing a subsample of them as a test group, while the remaining folds are used for training the model. Then, with the model calibrated with the training folds, the model predictions are obtained and compared with the test data in order to obtain the prediction error. This procedure is repeated over and over with all possible combinations of training and test folds, keeping track of the prediction errors. The model is thus chosen, as well as its parameters, that reduce the prediction error to its minimum value in all possible combinations. A prediction model that optimizes the out-of-sample performance is thereby obtained (Tiffin, 2016).

Gradient Boosting Machine

The gradient boosting machine technique involves building a set of decision trees in which the idea is to accumulate and train "weak prediction models"[11] so as to obtain an accurate prediction of the variable of interest by means of an ensemble model with heterogeneous weights. The prediction models are trained using the errors of the model representing the ensemble of weak prediction trees up to a given time (Natekin and Knoll, 2013). In this sense, the subsequent models make it possible to classify observations that previous models did not classify correctly. Consequently, weak prediction models successively improve the regression performance relative to that of the previous model (stepwise progressive approach).[12] The prediction made by each aggregate tree is thus accumulated, together with the sequence of predictions of the previous trees, in an effort to correct or improve the final model output. Following Boehmke and Greenwell (2020), the gradient boosting machine methodology can be summed up in the following steps:

1. Fit a regression with a regression tree $y = F_1(x)$
2. Obtain the residuals from this tree and then use them to fit the next tree $F_2(x) = F_1(x) + e_1$
3. Similarly, obtain the residuals of tree $F_2(x)$ and fit the next tree $F_3(x) = F_2(x) + e_2$
4. The process continues until a cross-validation metric stops it.

Following this process Z number of times (Z number of trees) we obtain a stagewise additive model expressed by equation 2:

$$F(x) = \sum_{z=1}^{Z} F_z(x) \qquad (2)$$

In the foregoing expression, $F(x)$ is a regression tree, which is successively updated by adding weak prediction models. Then we want to minimize the following expression by fitting $F_z(x)_z$ with new regression trees:

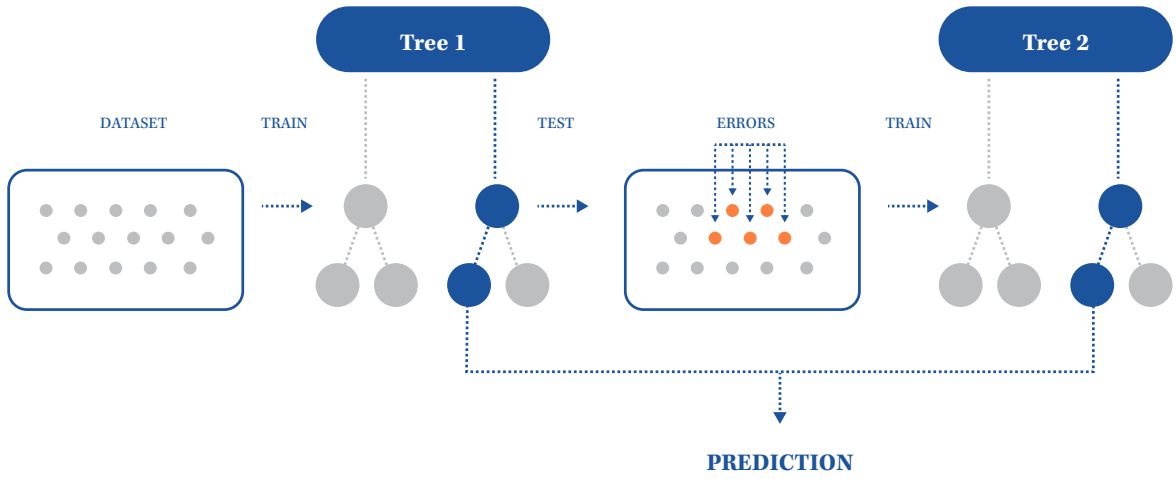
$$L = \sum_{z} L\big(y_z, F_z(x)\big) \qquad (3)$$

11 A weak prediction model is one whose performance is not significantly better than the performance that could be obtained by chance (Brownlee, 2016). The use of weak prediction models has some advantages over accurate prediction models: it is not computationally demanding, the models learn slowly on the basis of previous ensembles, and it helps avoid the common problem of overfitting.

12 In gradient boosting machine, "shortcomings" are identified by a gradient, and that gradient tells us how to improve the model (Sick, 2018).

By adding these weak regression trees, we can devise a powerful model that can produce an accurate quarterly GDP forecast. Figure 2 offers a simplified illustration of the mechanism behind gradient boosting machines.

FIGURE 2. SIMPLIFIED REPRESENTATION OF THE GRADIENT BOOSTING MACHINE

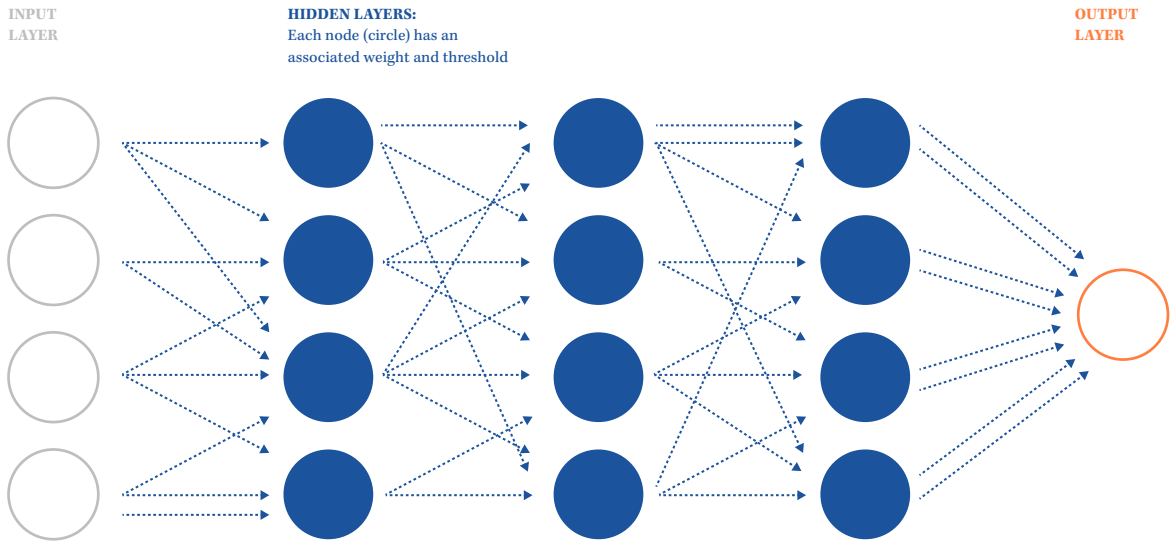


Source: Adapted from Boehmke and Greenwell (2020).

Neural networks

Neural networks consist of a group of nodes (which can be thought of as linear regressions) that are connected to each other in different layers (Mitchell, 2010). In simple terms, each input is assigned a weight to measure the importance of each variable, and the inputs are then multiplied by the corresponding weight to yield a result. The result, in turn, is transmitted to an activation function (stepwise) that determines if the output has exceeded the corresponding threshold to activate the data and move to the following layer/nodes. This process is repeated in each layer of the neural network, and thus the information contained in the previous nodes serves to feed the following layers/nodes (see Figure 3).

FIGURE 3. SIMPLIFIED REPRESENTATION OF ARTIFICIAL NEURAL NETWORKS



Source: Adapted from IBM Cloud Education (2020).

Training neural networks involves using a loss function to be minimized. To perform this minimization, the standard procedure is to use the stochastic gradient descent algorithm. This algorithm helps locate the necessary partial derivatives with respect to the weights and biases, by means of an update function. It is possible to update the weights and biases until we minimize the loss function, and produce an accurate prediction of the outcome variable equation 4 provides a simple representation. Figure 3 presents a simplified descriptive diagram of the mechanism behind this algorithm.

$$y = F\left(\sum_{i=1}^{m} w_i x_i + bias\right) \quad (4)$$

Ridge regression

Ridge regressions are among the supervised machine learning algorithms fitted through penalized linear regressions. The idea behind this method is to reduce the complexity of a regression with a large number of explanatory variables. The method thus entails a trade-off between the bias of the coefficients and the variance (Boehmke and Greenwell, 2020). Figure 4 illustrates this bias-variance trade-off. Ordinary least squares (OLS) seek to reduce the bias as much as possible, whereas penalized regressions accept an increase in the bias so as to reduce the variance of the coefficients and thus find the model's optimal complexity. Basically, ridge regressions penalize the predictors according to their size by shrinking them to values that tend toward zero. The final size of each coefficient will depend on its importance in explaining the target variable. The ridge regression seeks to minimize a loss function (similar to that minimized in OLS, but by adding the term containing the regularization coefficient (λ)) as in expression 5. Expression 6 shows how the coefficients are obtained by this method.

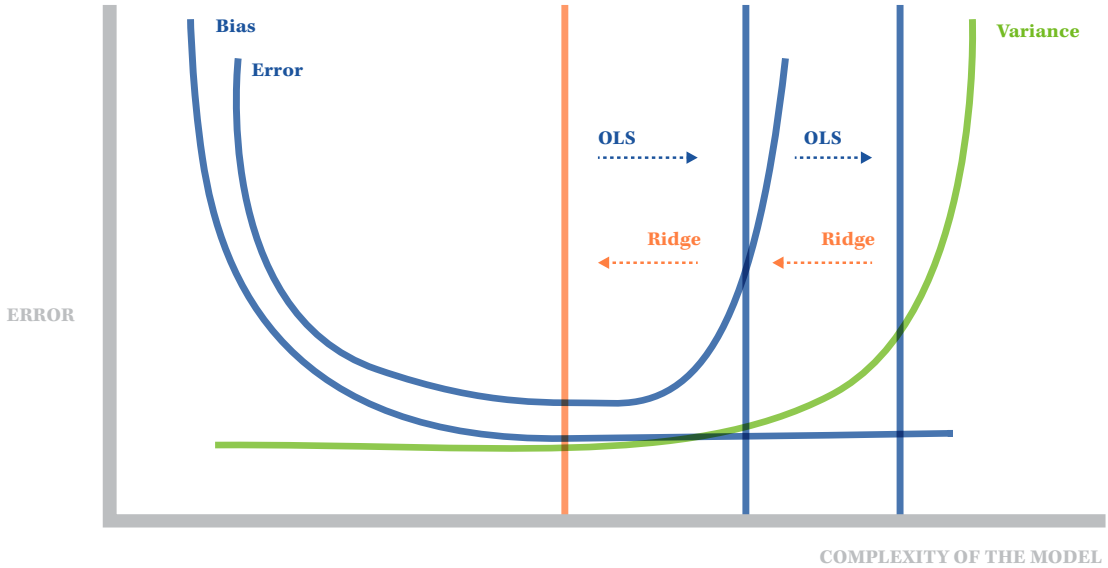
$$L(\hat{\beta}) = \sum_{i=1}^{n} \left(y_i - x_i'\hat{\beta}\right)^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 \quad (5)$$

$$\hat{\beta}_{ridge} = \left(X'X + \lambda I\right)^{-1}\left(X'Y\right) \quad (6)$$

In expressions 5 and 6, λ is defined as the regularization (or penalty) coefficient. When parameter λ takes

the value of zero, the estimated coefficients come from OLS. It is also important to note that when the regularization coefficient increases, the variance tends to decline, albeit at the cost of an increase in the bias. In this sense, the selection of the regularization parameter is not trivial. For that purpose we can use an information criterion (IC) or perform a cross-validation and select the parameter value that minimizes this metric, such as the mean square error (MSE) or root mean square error (RMSE). The latter procedure is the one used in this paper.

FIGURE 4. ILLUSTRATION OF THE BIAS-VARIANCE TRADE-OFF, OLS VS RIDGE



Source: Prepared by the authors.

Lasso regression

Lasso (least absolute shrinkage and selection operator) regressions share the same principle as ridge regressions and are among the regularized regressions. The difference between the two is the penalty coefficient. Specifically, in lasso regressions the penalty is given by the second term of expression 7 (loss function):

$$L(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} |\hat{\beta}_j| \qquad (7)$$

As in the ridge regression, the optimal value of λ can be obtained by applying an IC or a cross-validation and selecting the λ that produces the lowest MSE or RMSE. One of the main differences between the lasso and ridge regressions is that the lasso regression, given the kind of penalty it applies, allows for a selection of variables (for variables with less explanatory power, the coefficient is zero); this is not the case with the ridge regression penalty. In this regard, the lasso regression allows us to perform a feature selection which discloses those variables that, according to the regression, have a statistically significant relationship with quarterly GDP.[13]

13 As is evident below, for Belize a preselection of variables based on the lasso regression was used to increase the predictive capacity of the models in a context marked by significant challenges in the availability of statistics.

Elastic net regression

As may be intuited, one of the main limitations of the lasso regression is that the selection of variables that have greater power to predict the target variable depends significantly on the dimensionality of the variables in the database of interest, as well as on the availability of information. Hence the efficient selection of variables is conditioned by the dimensionality and availability of information. With ridge regressions, one of their main limitations is that they are highly dependent on regressors being significantly correlated with the regressand; if they do not, there is a risk that the prediction's efficiency might be affected. One way of resolving these limitations is to combine the methods described above (Boehmke and Greenwell, 2020)—that is, penalization of the coefficients is given by a combination of the penalties in the lasso and ridge regressions. Thus we have expression 8. In the elastic net regression, in addition to finding an optimal value for parameter λ, we also have to find a value for α. This latter parameter is a weight with a range between zeroo (ridge regression) and 1 (lasso regression).

$$L(\hat{\beta}) = \frac{\sum_{i=1}^{n}\left(y_i - x_i'\hat{\beta}\right)^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}\left|\hat{\beta}_j\right|\right) \quad (8)$$

nowcasting techniques offer an effective tool to fill the gap arising from the lag in the publication of macroeconomic indicators

Belize and
El Salvador

Data

Collecting data to conduct a nowcasting exercise is a process that poses significant challenges related directly to a country's statistical capacity. Generally, these exercises call for a wide array of variables that are associated to some degree with movements in quarterly GDP. Another important consideration is that the exercise improves as the time series of the predictor variables are fully available, so there has to be a trade-off between the timeliness and efficiency of the forecasting models. This makes sense when we consider that the predictor variables used are published at different intervals, including monthly, bimonthly, and quarterly. The data used in the two countries are described below..

Belize

The goal of the exercise for Belize is to nowcast seasonally adjusted quarterly GDP (constant 2000 prices, millions of Belize dollars). For the purposes of this exercise, information was collected on more than 108 variables.[14] The database was built by collecting information from several sources, including the Central Bank of Belize (CBB), the Statistical Institute of Belize (SIB), the Belize Tourism Board (BTB), and US Federal Reserve Economic Data (FED-FRED). The variables are grouped into real sector variables, fiscal variables, external sector variables, monetary variables, and variables from other countries that are related to movements in Belize's quarterly output.[15] [16] The frequency of the variables is mixed, while the frequency of GDP is quarterly; most of the variables are monthly. The database covers the period 1994 to 2021. Some of these monthly variables are available at the end of the corresponding quarter. That circumstance allows for a preliminary forecast about 12 weeks before the SIB's publication of the official quarterly GDP figure, and a final forecast four weeks before the publication.

All monetary variables expressed in dollars were deflated with the general CPI. All variables were standardized to quarterly values. To that end, monthly variables were divided into flow variables, index/rate variables, and stock variables. The sum of the months making up each quarter is used for the flow variables (for instance, this is the case for exports and imports).[17] For the second and third group of variables, the average value of the months corresponding to each quarter was taken.[18] Annex 1 presents a list of the variables included in this exercise and their corresponding classification. To feed the models, we used the quarterly year-on-year growth rate of the variables to cope with the non-stationarity that normally affects macroeconomic time series, and because the Bank's interest centers on measuring movements in quarterly GDP growth.[19]

14 The selection of variables was based on Aasaavari et al. (2018), and on discussions between the consultant and the IDB team overseeing the initiative. However, after applying the lasso regression and feature selection, the number of variables was reduced to 24.

15 In this particular case, monthly information was collected on real-sector indicators from the United Kingdom and the United States, two of Belize's main trading partners.

16 As mentioned earlier, some variables that can be expected to be strongly related to movements in quarterly GDP could not be obtained, or their publication intervals are not conducive to their inclusion in nowcasting exercises. In cases such as monthly sectoral output or fiscal statistics (tax revenues, for example), monthly data is published at intervals that seriously constrain the availability of this information for nowcasting purposes. Similarly, there is no public access to monthly data on the number of people employed in the formal sector of the economy, nor monthly information on energy production and consumption. In nowcasting exercises, these variables generally have a strong association with quarterly GDP.

17 In cases where information is only available for one month in the quarter, it is assumed that the variable will have the same value in the other two months of that quarter. When two months are available, it is assumed that the value of the variable in the third month will be the same as in the second month. This is regarded as a simple procedure that does not add more statistical noise to the information, and does not pose a significant risk to the statistical capacity of the calibrated models. Nonetheless, autoregressive or other models can be used to complement the information for each quarter.

18 In cases where information is available for only one month in the quarter, that value is taken as the average for the quarter. In cases where information is available for two months, the average of the two months is taken as the average for the quarter.

19 As part of the exercise, the models were also fed with the level variables, yielding acceptable forecasts and efficiency. This puts into perspective the fact that machine learning tools provide a more flexible framework than standard time series models.

El Salvador

The goal of the exercise for El Salvador is to predict the chained volume index (2014 benchmark) of quarterly GDP for the seasonally adjusted series published by the Central Reserve Bank (BCR). For this exercise, a primary database was built with 98 macroeconomic variables that were collected on a monthly or quarterly basis from January 2005 up to (depending on the variable) March 2021.[20] Most of these variables were obtained from the BCR, although we also used data from the Salvadoran Social Security Institute (ISSS), US Federal Reserve Economic Data (FED-FRED), and information on sales and investment volume collected by the Salvadoran Foundation for Economic and Social Development (FUSADES).

Variables expressed in nominal dollars were deflated with the CPI, or the WPI (Wholesale Price Index)/PPI (Producer Price Index) for variables related to foreign trade. The frequency of the variables is mixed; while the chained volume index of GDP is quarterly, most of the variables are monthly. All the variables, therefore, were standardized to quarterly variables. Monthly variables were divided into flow variables, index/rates variables, and stock variables. For the former, the sum of the months making up each quarter was taken. For the other two variables, the average value of the months corresponding to each quarter was taken (see Annex 2).[21] Finally, the models were fed with the year-on-year quarterly growth rate of the variables in the database, since this allows us to deal with the problem of non-stationarity while GDP growth is the variable of interest.[22]

20 The variables were chosen on the basis of Aasaavari et al. (2018) and discussions between the consultant and the IDB team overseeing the initiative.

21 This is a difference compared to Aasaavari et al. (2018), in which the stock variables are represented by the value of the final month of the quarter. In this paper, using the average of the stock variables is deemed to offer a better representation of the behavior of the stock variables in the quarter.

22 As part of the exercise, the models were also run with the variables in level, yielding an in-sample and out-of-sample fit very similar to that obtained with the quarter-on-quarter growth rates.


AI
TECHNOLOGY
DATA CENTER LA- 204.22.77.43
AUTONOMOUS
ALGORITHM
215
TB
ID-8A
ONLINE
SCANNING
FACIAL RECOGNITION
36
ON
CANCEL
BACKWARD CHAINING
INTERVAL DATA
HELP?


Luz Estrella
Apopa
ALCALDESA
2009-2012

Results

This section presents the results obtained from calibrating the six machine learning methods and their corresponding ensembles to forecast quarterly GDP for Belize and El Salvador.[23] The results are presented for penalized regressions and then for the other models.[24]

In the case of Belize, the results show that the penalized regression models produce a better fit than the other methods, as measured by the RMSE. Specifically, an ensemble of these regressions yields a forecast very close to the observed outcome between the second and fourth quarters of 2020. The model predicted a 22.9 percent fall in GDP in the second quarter (the actual figure was 23.9 percent); a 15.0 percent fall in the third quarter (the actual figure was 12.8 percent); and a 14.2 percent fall in the fourth quarter (the actual figure was 13.1 percent). For the first quarter of 2021 the model forecasts a 6.8 percent drop in GDP, while the actual decline was 8.4 percent.

In the case of El Salvador, the model with the greatest predictive efficiency is the ensemble of penalized regressions. For example, for the second quarter of 2020 the model predicted a fall of 17.8 percent (the actual figure was -19.4 percent); the model predicted an 8.0 percent contraction in the third quarter, while the figure published by the BCR was -9.9 percent. For the fourth quarter the model predicted a fall of -2.3 percent, and the actual figure was -2.3 percent. Finally, for the first quarter of 2021 the model forecast growth of 1.2 percent relative to the first quarter of 2020 (the actual figure was 3.0 percent). As mentioned earlier, El Salvador has greater statistical capacity than Belize. This is reflected in the fit of the models, because the RMSEs obtained for El Salvador are generally lower than those for Belize. The following subsections present the results and selection of the best model for Belize and El Salvador.

Penalized regressions

Table 1 and Figure 5 present the results of the following methods: lasso regression, ridge regression, and elastic net regression. The regressions were obtained for the growth-rate variables. As Figure 5 shows, the forecast yielded by the penalized regressions is quite good relative to movements in quarterly GDP in both countries. In general, the fit is better for El Salvador than for Belize, both in-sample and out-of-sample. As mentioned, this is because in the case of Belize, variables that are likely to be highly correlated with movements in quarterly GDP—such as electricity production and consumption, or monthly employment series—were not available. They were available for El Salvador. An important aspect of these models is that they can capture the sharp decline in output in the second quarter of 2020, the third quarter of 2020, and the fourth quarter of 2020 in both countries. This underscores the capacity of machine learning models to identify significant changes in the time series, as well as their ability to deal with existing nonlinearities in the data collected.[25] One of the advantages of machine learning techniques is that they emphasize out-of-sample performance in determining accuracy. In this particular case, the models were trained by dividing the data by 70–80 percent as part of the training, and using the cross-validation process. The training of the models included the period 2006-Q1 to 2020-Q2. The forecasts from these calibrations are compared with the test dataset, and the RMSE is estimated.

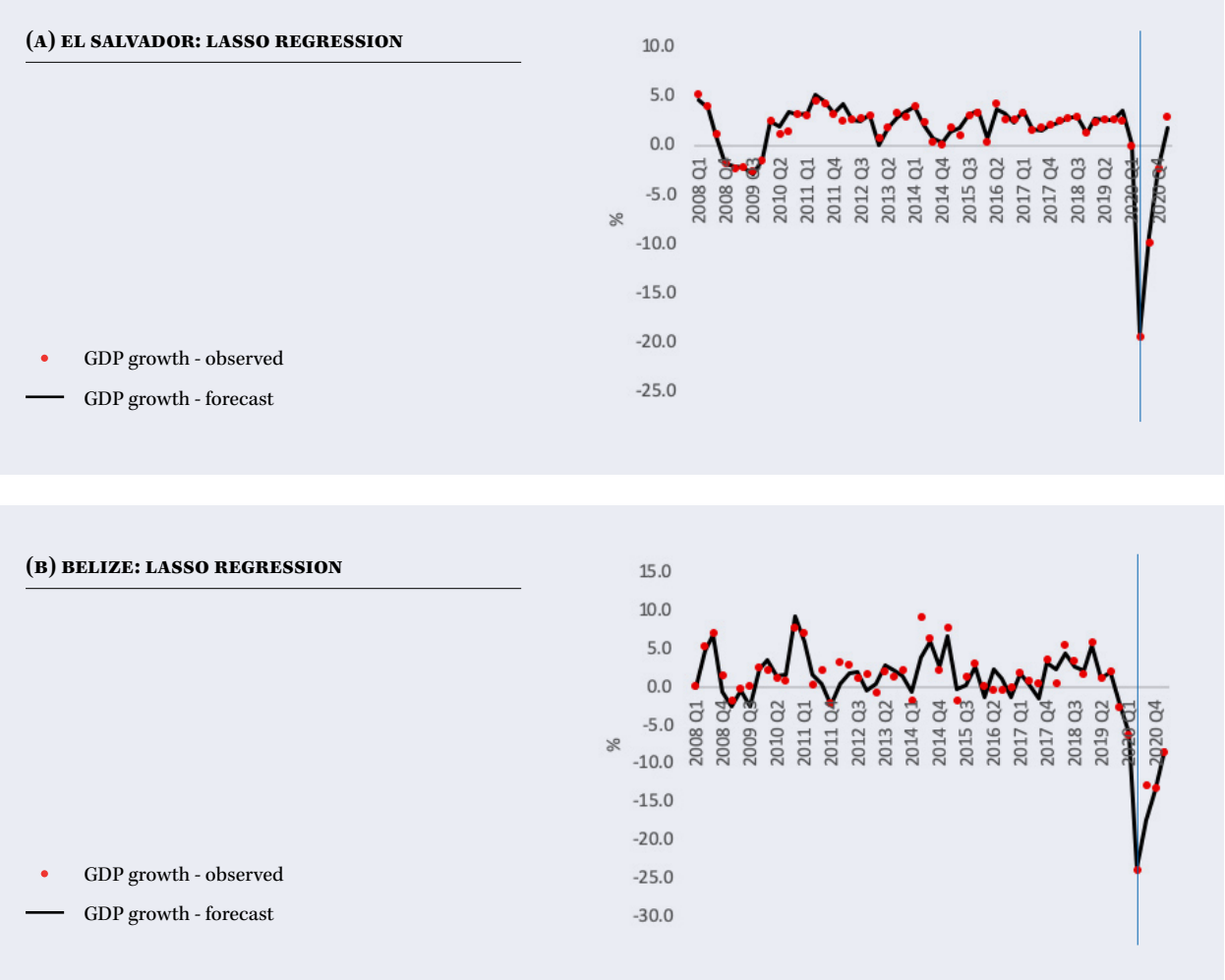
23 R software was used to obtain these results.

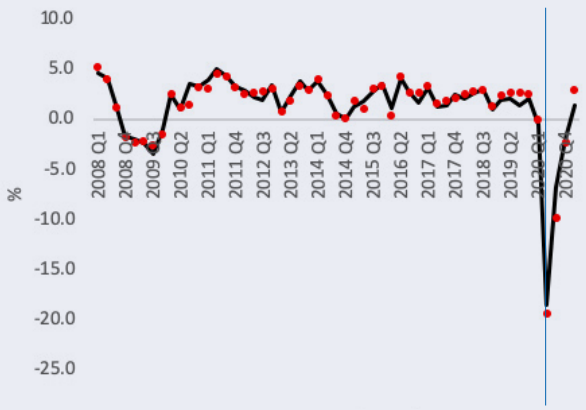
24 As mentioned above, the Belize models were calibrated from a preselection of variables using the lasso regression. In the Belize case, then, the models were fed with 24 variables selected by this penalized regression. This approach improved the in-sample and out-of-sample fit. Annex 3 presents the list of variables selected by the lasso regression for Belize. In the case of El Salvador, there was no need for a preselection of variables to calibrate the models. Nonetheless, Annex 4 presents the list of variables selected by the lasso regression for El Salvador.

25 This exercise also included the estimate of a dynamic factor model for El Salvador, and it was evident that the model had greater difficulties in predicting falls in output in the second, third, and fourth quarters. Moreover, the fitted machine learning models yielded a lower prediction error.

Table 1 shows the results of the MSE and RMSE for the penalized regressions. The RMSE is measured in the units of the outcome variable (quarterly GDP growth in the two countries). The three models produce a similar RMSE for both countries. However, the lasso regression (for El Salvador) and the ridge regression (for Belize) are the models that reduce the RMSE to a minimum, indicating the best out-of-sample fit. In comparative terms, the data in Table 1 show that the models predict a better fit in the case of El Salvador because the RMSE is lower in all cases, confirming what was said earlier about the availability of information. The models for both countries were trained using data for the period Q1-2006/Q2-2020. Hence the forecasts produced after Q2-2020 (to the right of the blue line) act as a true-out-of-sample test to measure the predictive capacity of the models. As can be seen from the figure, in all cases the models yield results very close to the growth rates for the third and fourth quarters of 2020 and the first quarter of 2021, thus confirming the models' predictive capacity.

FIGURE 5. RESULTS OF PENALIZED REGRESSIONS FOR EL SALVADOR AND BELIZE

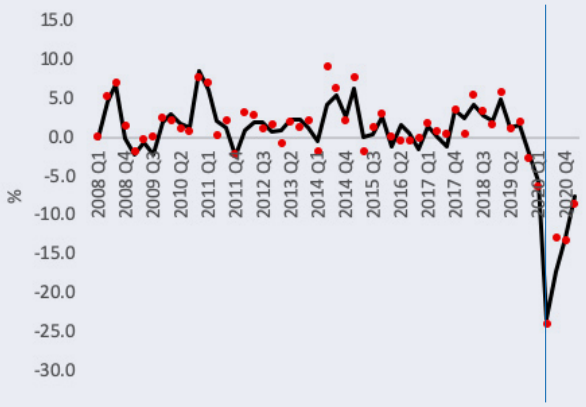




(C) EL SALVADOR: RIDGE REGRESSION

GDP growth - observed

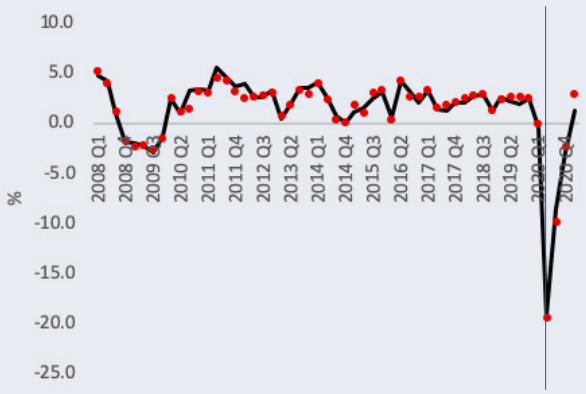
GDP growth - forecast



(D) BELIZE: RIDGE REGRESSION

GDP growth - observed

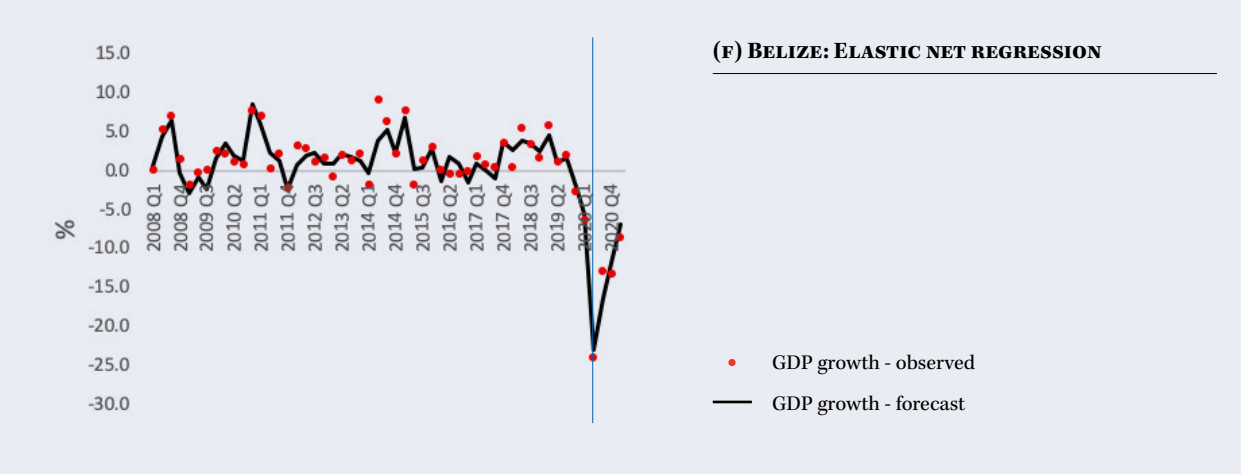
GDP growth - forecast



(E) EL SALVADOR: ELASTIC NET REGRESSION

GDP growth - observed

GDP growth - forecast



Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

TABLE 1. MSE AND RMSE OF PENALIZED REGRESSIONS

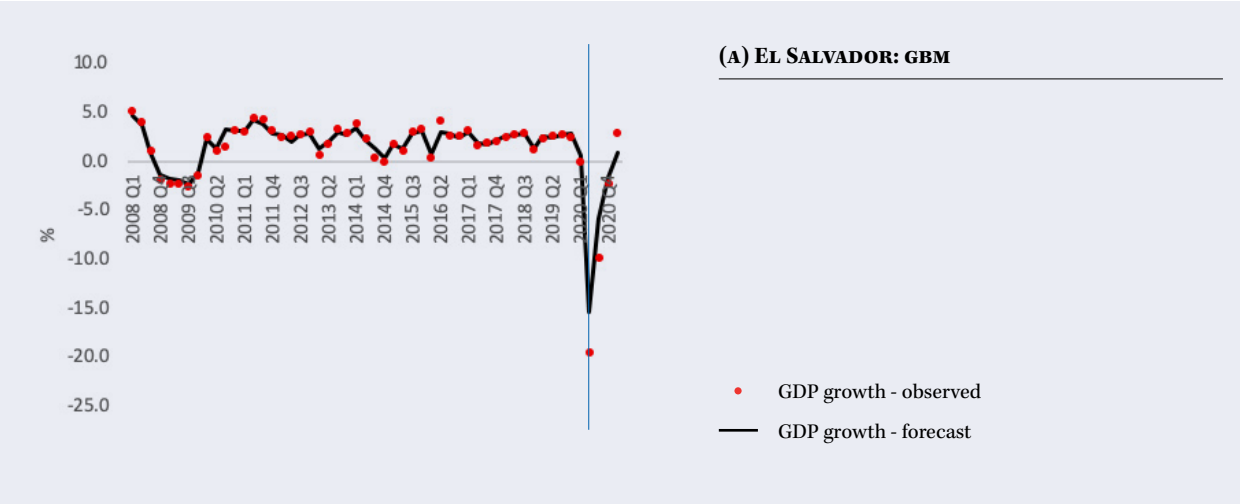
Model	El Salvador		Belize	
	MSE	RMSE	MSE	RMSE
Lasso regression	0.60	0.77	4.89	2.21
Ridge regression	0.66	0.81	3.16	1.78
Enet regression	0.60	0.78	3.75	1.94

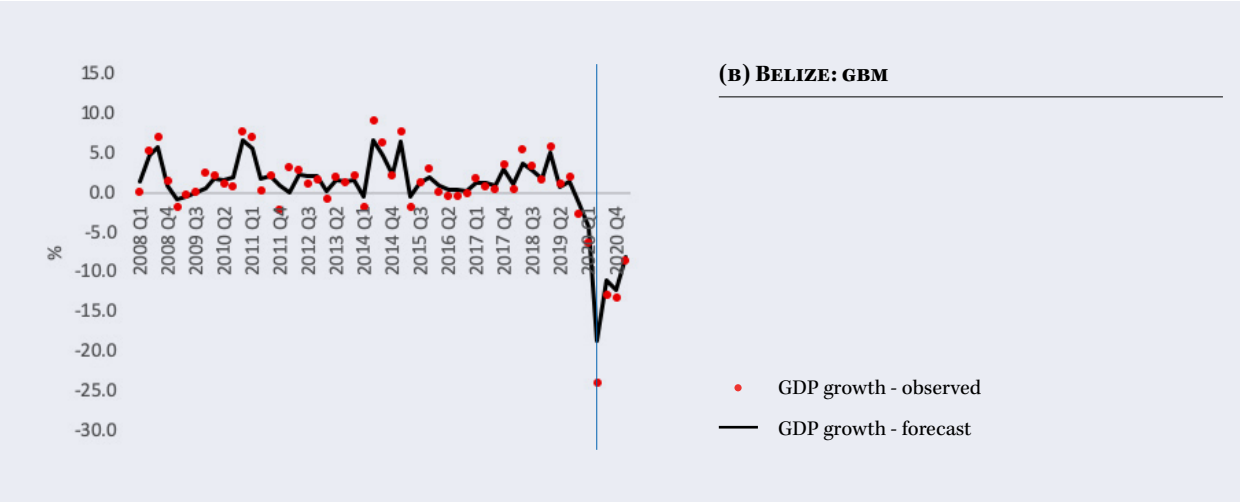
Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

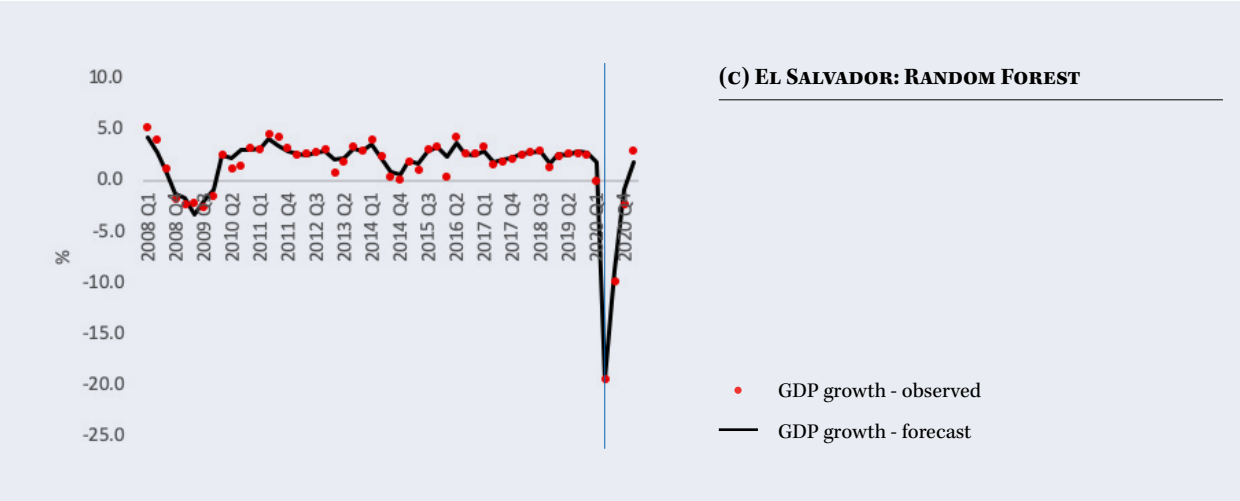
Random forest, gradient boosting machine, and neural network

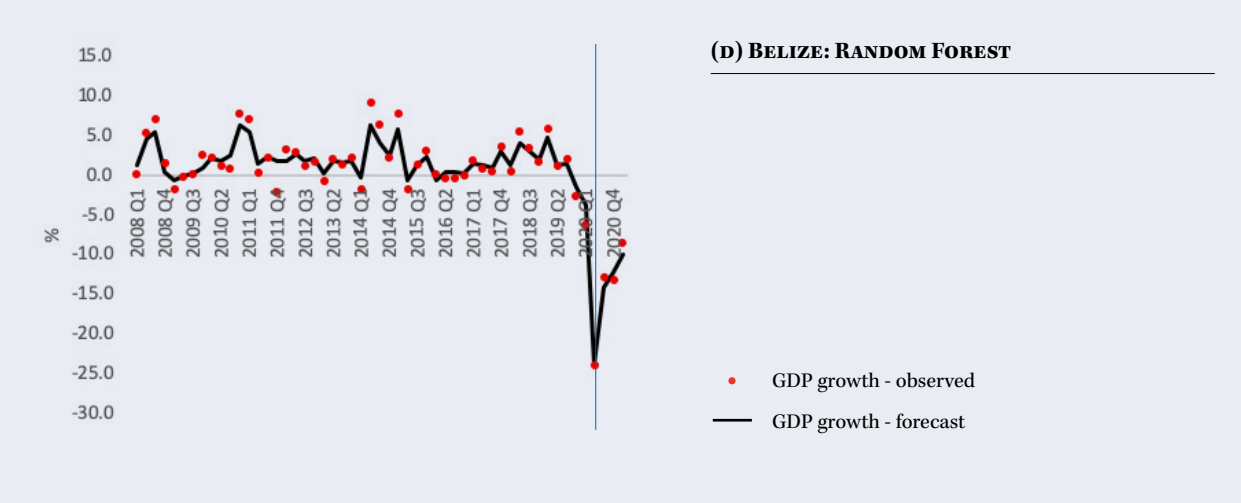
In addition to the penalized regressions presented in the previous section, the following models were also estimated: random forest, gradient boosting machine, and neural networks (see Table 2 and Figure 6). The models were calibrated for the growth-rate variables. Like the penalized regressions, these models are able to capture the sharp fall in GDP in the last three quarters of 2020. Figure 6 shows that the fit of the in-sample models is accurate. It is also evident, once again, that the data have a better fit in the case of El Salvador than in the case of Belize. In particular, the random forest and GBM models have a good in-sample and out-of-sample fit, which is represented by the dots to the right of the blue line (true-out-of-sample). In general, the models perform well in predicting GDP declines in the last three quarters of 2020 and the first quarter of 2021, with the exception of GBM, which in both countries fails to capture the abrupt drop in output in the second quarter of 2020. As mentioned earlier, machine learning techniques emphasize out-of-sample performance, and with that in mind the models were trained with information covering Q1 2006 to Q2 2020. With this information we obtain the predictions and compare them with the test data (excluded from the training), so as to secure the RMSE of each model. Table 2 presents these results. Of the models calibrated for El Salvador, GBM yields the lowest RMSE, followed by neural network and random forest. For Belize, the models producing the lowest RMSE are GBM and random forest.

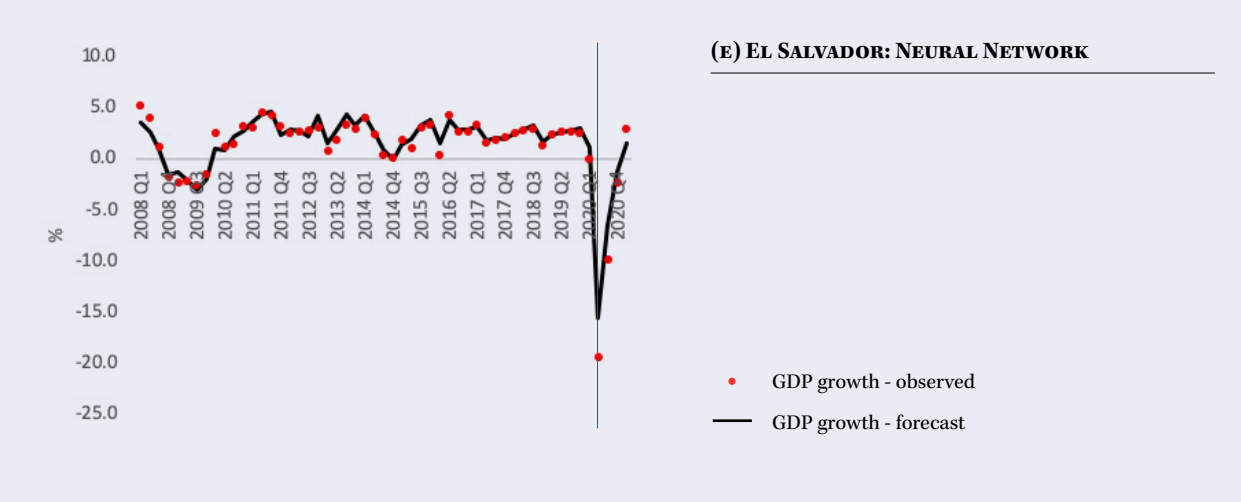
FIGURE 6. GBM, RANDOM FOREST AND NEURAL NETWORK RESULTS FOR EL SALVADOR AND BELIZE

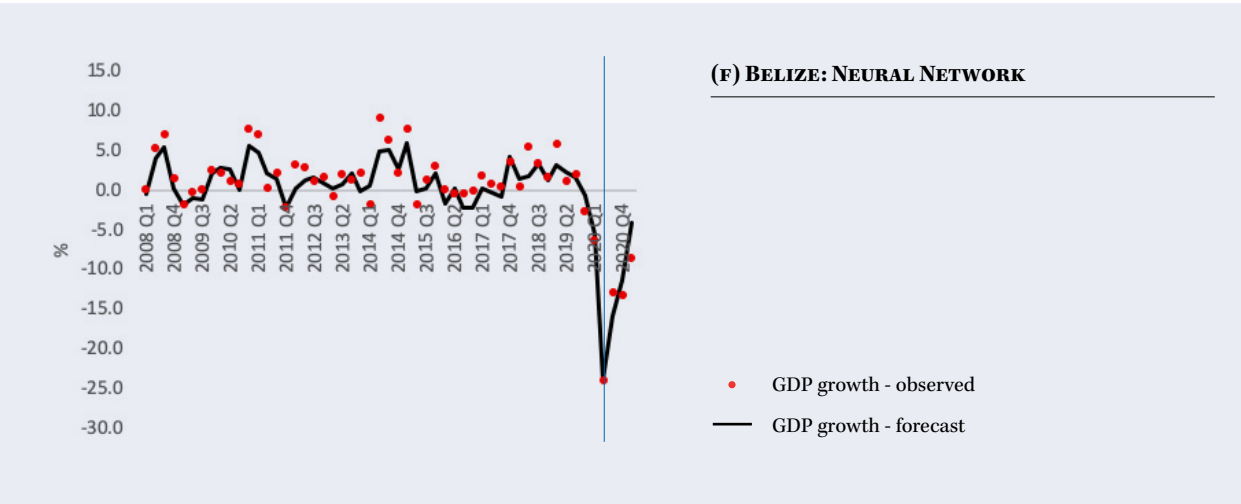












Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

TABLE 2. MSE AND RMSE OF GBM, NNET, AND RANDOM FOREST

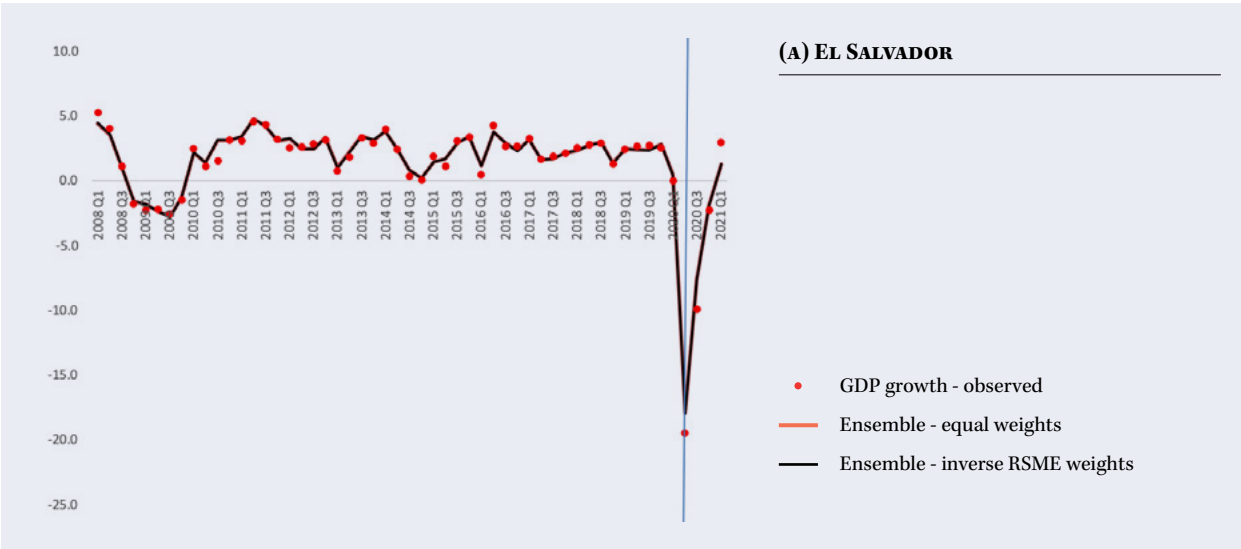
Model	El Salvador		Belize	
	MSE	RMSE	MSE	RMSE
Gradient Boosting Machine	0.67	0.82	3.33	1.83
Neural Network	0.95	0.98	3.56	1.89
Random Forest	1.00	1.00	3.35	1.83

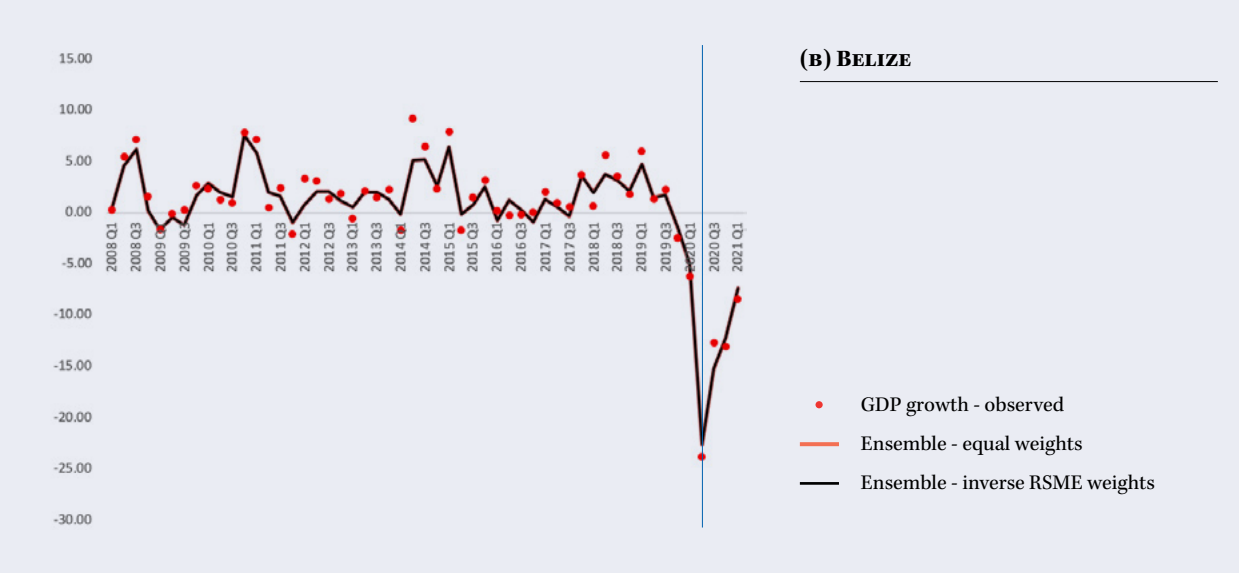
Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

Ensemble approach

One of the advantages of nowcasting exercises is that forecasts from different models can be combined to exploit the particularities of each of them. This is known as the ensemble approach. This subsection presents the results of this exercise with the six models calibrated for the two countries. There are different means of undertaking this combination, and two of them are used in this paper: simple average (equivalent weights) and weights devised with the inverse of the RMSE. The former method uses all the forecasts produced by all the models to obtain an average of all the forecasts. In the other method, a weight is devised on the basis of the inverse of the RMSE produced by each model. Figure 7 shows the results of this exercise for the two countries. To verify the impact on reducing the prediction error, Table 3 presents the RMSE estimate for both ensembles. Table 3 shows clearly that the prediction error has been significantly reduced for both countries. This indicates that combining the forecasts from the machine learning models used in this paper yields greater predictive efficiency relative to the models taken individually, as demonstrated by the RMSE metric.

FIGURE 7. ENSEMBLE APPROACH FOR EL SALVADOR AND BELIZE





Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

TABLE 3. RMSE AND MSE OF ENSEMBLES FOR EL SALVADOR AND BELIZE

Model	El Salvador		Belize	
	MSE	RMSE	MSE	RMSE
Ensemble – equal weights	0.29	0.54	1.45	1.21
Ensemble – inverse RMSE weights	0.29	0.54	1.44	1.20

Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

Selecting the best model

In this section we combine the results obtained in the previous sections, with a view to comparing the models and choosing those that minimize the RMSE and yield the best true-out-of-sample fit. Table 4 shows the comparison of the metrics of the six models for the two countries, as well as their corresponding ensembles. In individual terms, the results indicate that the penalized regression models produce a better fit than the other methods, since they yield comparatively lower RMSEs and can predict the second-quarter fall in GDP more accurately. These results are in line with previous findings by Tiffin (2016), who calibrated different machine learning methods to predict Lebanon's quarterly GDP. He found that the penalized elastic net regression produced the lowest RMSE and a better fit. For El Salvador, then, the lasso regression produces the lowest RMSE, whereas for Belize it is the ridge regression. The results also show, however, that predictive efficiency increases when the ensembles of the six models are used, since they have the lowest value for the RMSE metric. To test the predictive power of these models, we proceeded to train the models with information up to the last quarter of 2019, so that the forecasts for all quarters of 2020 and the first quarter of 2021 are true-out-of-sample. Figures 8 and 9[26] and Table 5 present the results of these models (fewer years are presented in order to observe the fit of the models more clearly).

26 The ensembles presented in Figure 9 and Table 5 were obtained only with the penalized regressions, since these yield the lowest RMSE.

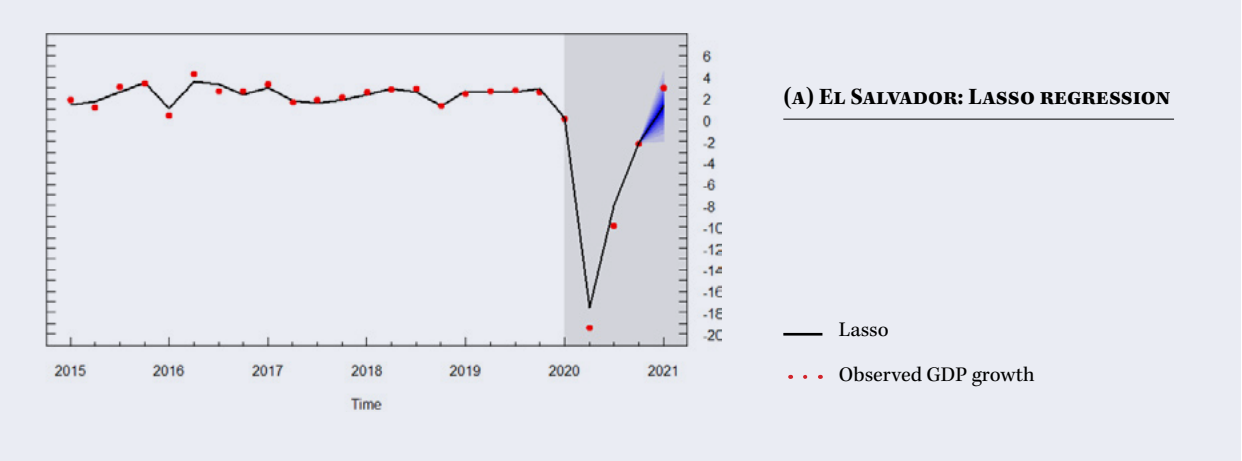
Table 4. MSE and RMSE of Penalized Regressions, GBM, NNET, Random Forest, and Ensembles

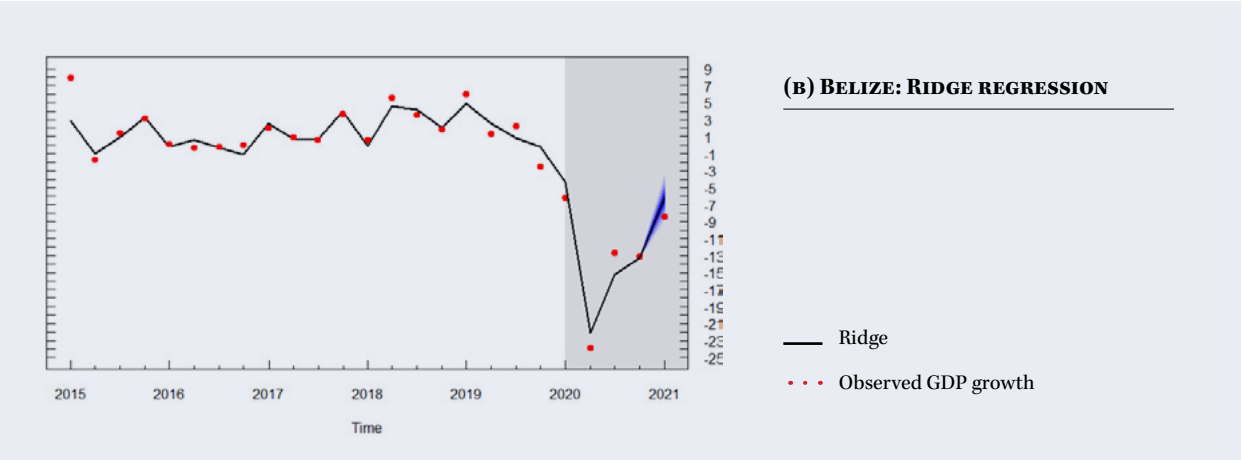
Model	El Salvador		Belize	
	MSE	RMSE	MSE	RMSE
Lasso regression	0.60	0.77	4.89	2.21
Ridge regression	0.66	0.81	3.16	1.78
Enet regression	0.60	0.78	3.75	1.94
Gradient boosting machine	0.67	0.82	3.33	1.83
Neural network	0.95	0.98	3.56	1.89
Random forest	1.00	1.00	3.35	1.83
Ensemble – equal weights	0.29	0.54	1.45	1.21
Ensemble – inverse RMSE weights	0.29	0.54	1.44	1.20

Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

Figures 8 and 9 indicate that the models produce a good in-sample fit, but also a good out-of-sample fit. Specifically, the forecasts in the gray area display a high level of accuracy in all quarters of 2020 for El Salvador. In the case of Belize, the model presented difficulties in predicting the decline in the second quarter of 2020. Essentially, this reflects something that has been pointed out throughout this paper: a country's statistical capacity is vitally important in ensuring accurate nowcasts. Despite the challenges posed by Belize's statistical capacity, the models do a good job of accurately approximating falls in output during the other quarters. This is another advantage of machine learning models, since the cross-validation process ensures the best fit even in exercises that pose significant challenges of statistical capacity. Table 5 shows the true-out-of-sample forecasts produced by the penalized regressions and their corresponding ensembles (which for both countries produce the lowest RMSEs). The models for the two countries yield fairly accurate estimates of quarterly GDP in each of them during 2020. The models demonstrate that they can recognize unprecedented falls in output relatively accurately, which again indicates the capacity of machine learning models to produce accurate nowcasts of output in a context of high uncertainty.

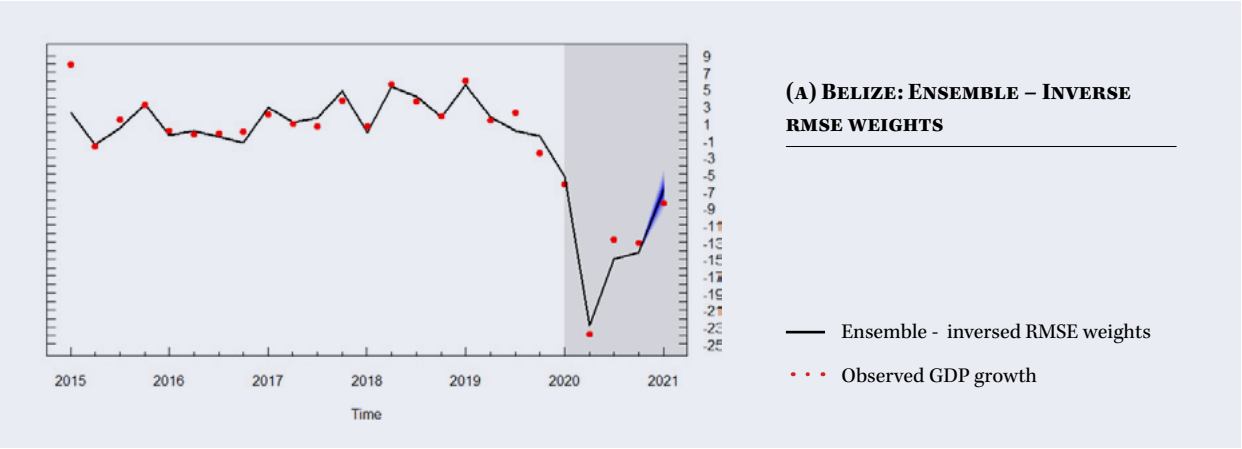
FIGURE 8. BEST-FIT PENALIZED REGRESSIONS FOR EL SALVADOR AND BELIZE

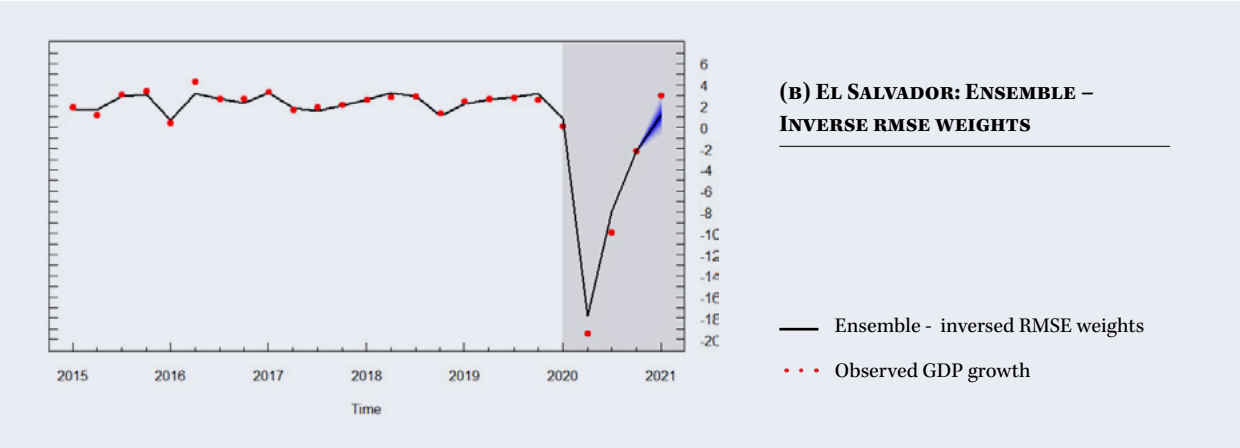




Note: the blue area forecasting Q1 of 2021 was devised using an adaptation of the fan charts methodology developed and implemented by the Bank of England.
Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

FIGURE 9. BEST-FIT ENSEMBLES FOR EL SALVADOR AND BELIZE





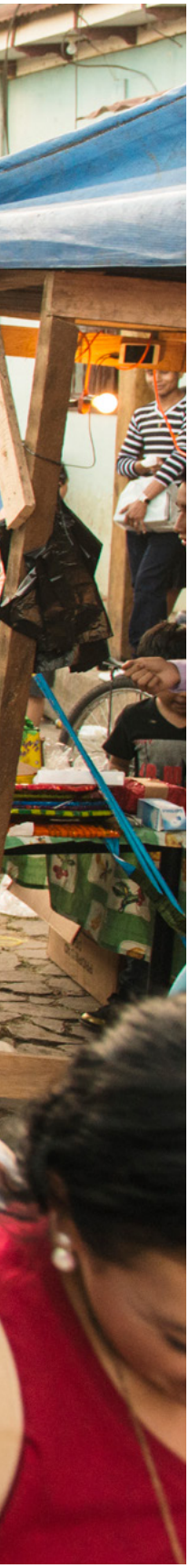
Note: the blue area forecasting Q1 of 2021 was devised using the an adaptation of the fan charts methodology developed and implemented by the Bank of England.
Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.

TABLE 5. TRUE-OUT-OF-SAMPLE FORECASTS

Period	El Salvador (%)					Belize (%)				
	Observed	1. Elastic net	2. Lasso	3. Ridge	4. Ensemble - inverse RMSE weights	Observed	1. Ridge	2. Elastic net	3. Lasso	4. Ensemble - inverse RMSE weights
2020 Q1	0.1	0.5	0.2	1.7	0.8	-6.2	-4.3	-5.8	-5.7	-5.2
2020 Q2	-19.4	-18.1	-17.5	-17.9	-17.8	-23.9	-22.1	-23.5	-23.2	-22.9
2020 Q3	-9.9	-8.3	-8.0	-7.6	-8.0	-12.7	-15.3	-15.0	-14.7	-15.0
2020 Q4	-2.3	-2.6	-2.2	-2.0	-2.3	-13.1	-13.3	-15.0	-14.6	-14.2
2021 Q1	3.0	1.0	1.3	1.2	1.2	-8.4	-6.1	-7.3	-7.0	-6.8

Source: Authors' estimate on the basis of BCR, ISSS, FUSADES, FED-FRED, CBB, SIB, and BTB.


LA FERIA
IMACASA



Institutional support

To a large extent, the exercise in Belize was carried out thanks to support from the Central Bank of Belize (CBB), which provided access to sectoral output statistics that improved the models' predictive capacity. At the same time, the IDB has benefited from the feedback it received from the CBB and SIB on preliminary presentations of this initiative.

In the case of Belize, a key aspect of the nowcasting project has been the complementarity of the work carried out by the IDB and the CBB. It should be noted that the CBB has its own nowcasting models: see, for example, Arana (2015). These models are expected to support the CBB's efforts to monitor and predict the performance of the economy. In addition to the ongoing dialogue with the CBB and SIB, the IDB project will provide the authorities with training, so as to build the institutional capacity to design and use nowcasting models and support their implementation at the institutional level by the CBB.

Efforts have been made to develop nowcasting tools in El Salvador. In 2009 the BCR, with support from the Economic Commission for Latin America and the Caribbean (ECLAC), developed a nowcasting model based on Kalman filter techniques, which considered a dynamic-factor model. The BCR model took into account 17 monthly variables, and made it possible to obtain an indicator consistent with a short-term forecast for aggregate GDP but not for its components. The interpretation was therefore limited to determining the direction of the effect.

The IMF subsequently developed a nowcasting model for the Salvadoran economy that took a disaggregated approach, to complement the one used by the BCR (Aasaavari, 2018), with a view to facilitating interpretation and providing an explanation of estimated economic dynamics. For the methodology, a bridge model analyzed from the production side was considered. The prime motivation was to complement the work being carried out on such models with innovative machine learning tools, which will be useful not only for El Salvador's statistical agencies but also for the IDB's work of constant monitoring and assessment of macroeconomic conditions in its member countries.

Concluding thoughts

As evidenced throughout this paper, the development of machine learning is a good option for nowcasting quarterly GDP. These tools offer significant advantages, such as emphasizing out-of-sample performance, detecting nonlinearities in the data and, in conjunction with that, modeling complex relationships among the variables of interest and predictors.

To get the most out of nowcasting models, it is crucial to have timely and high-frequency statistical information that is highly correlated with movements in quarterly GDP. This is why the calibration of the models in El Salvador and Belize revealed areas of opportunity for future exercises in forecasting economic activity. Specifically, the exercise made it evident that Belize has some challenges as regards the availability of information. For example, there was no access to information on some key variables such as energy production and consumption, as well as labor-market data. With respect to the timely publication of information, variables such as sectoral output and public-sector revenues and expenditures are published with lags that can significantly affect these exercises. Similarly, in both countries it is important that the publication of preliminary GDP figures and other variables be as accurate as possible, since substantial revisions to them can undermine the reliability of the exercise. Finally, for future exercises, it is worth noting the importance of collaboration and feedback with the statistical authorities. Specifically, the feedback received in preliminary presentations of this exercise to the Central Bank and the Statistical Institute of Belize made it possible to strengthen the initiative.

Finally, the calibration of nowcasting models is a dynamic process that is refined over time. Hence the importance of continuing to implement them once the initial phase has been completed, because the series of steps involved in these exercises is an intensive process. As mentioned earlier, the first task is to choose the nowcasting models that best fit the variables to be predicted. In our case, there were six models among which options from the machine learning range were selected. These models were fed with national data produced by the countries, as well as data on the international contexts drawn up by other institutions such as the US Federal Reserve. Important in this step is the thoroughness with which the data are selected and handled, as well as the ways in which the data are transformed, as needed. This is the most important task. It depends on the quality of the statistics that are taken into account, and on astuteness in selecting the variables that best explain the behavior of each economy. For each economy, both the model with the best fit and the variables that best explain its economic performance vary. This is where one of the main added values of this estimation exercise becomes apparent.

In summary, the benefits of these tools are substantial. The predictions have been taken into consideration in the macroeconomic analyses and assessments that the IDB carries out for both economies, and in Belize they have been of resounding benefit for the Central Bank. Given the nature of the exercise and the benefits of this kind of tool, it is highly recommended for the other countries of the CID region, since it can model economic growth for different types of economies almost immediately and with a high degree of accuracy.

Bibliography

Aasaavari, N., Di Vittorio, F., Giri, R., Lariau, A., Lissovolik, B., Teodoru, I. (2018). El Salvador, Selected issues. IMF Country Report No. 18/152. International Monetary Fund Washington, D.C.

Arana, Rumile (2015). Evaluating Short Term Bridge Model Forecasts of Real GDP in Belize: A Quarterly Approach. Research Department, Central Bank of Belize.

Banco Central de Reserva de El Salvador (2021). Base de Datos Económica. Retrieved January 2021, from https://www.bcr.gob.sv/bcrsite/?cat=1000&lang=es.

Belize Tourism Board (BTB) (2021). Belize tourism statistics. Retrieved January 2021, from https://belizetourismboard.org/belize-tourism/statistics/.

Board of Governors of the Federal Reserve System (US), 6-Month Treasury Bill: Secondary Market Rate [TB6MS]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TB6MS.

Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [FEDFUNDS]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FEDFUNDS.

Board of Governors of the Federal Reserve System (US), Industrial Production: Total Index [INDPRO]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/INDPRO.

Board of Governors of the Federal Reserve System (US), Industrial Production: Manufacturing (NAICS) [IPMAN]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/IPMAN.

Boehmke, B. and Greenwell, B. (2020). Hands-On Machine Learning with R. Taylor & Francis.

Bolhuis, M. and Rayner, B. (2020a). Deus ex Machina? A Framework for Macro Forecasting with Machine Learning. IMF Working Paper.

Bolhuis, M. and Rayner, B. (2020b). The More the Merrier? A Machine Learning Algorithm for Optimal Pooling of Panel Data. IMF Working Paper.

Breiman, L. (2001). Random Forest, Machine Learning, Vol. 45, pp. 5–32.

Breiman, L. and Cutler, A. (n.d.). Random Forests. Retrieved from https://www.stat.berkeley.edu/~breiman/RandomForests/.

Brownlee, J. (2020). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning, XGBoost. Retrieved from https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/.

Central Bank of Belize (CBB) (2021). Rates and statistics. Retrieved January 2021, from https://www.centralbank.org.bz/rates-statistics.

Chakure, A. (2019). Random Forest Regression along with its Implementation in Python. Retrieved from https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f.

DIGESTYC. (2020). Encuesta de Hogares y Propósitos Múltiples 2019.

DIGESTYC. (2021). Encuesta de Hogares y Propósitos Múltiples 2020.

FUSADES (2020). Encuestas de Confianza Empresarial.

IBM Cloud Education (2020). Neural Networks. Retrieved from https://www.ibm.com/cloud/learn/neural-networks.

IDB 2021. Inequality and social discontent: How to address them through public policy. Economic report on Central America, Haiti, Mexico, Panama and the Dominican Republic. Inter-American Development Bank.

IMF (2021). World Economic Outlook. January 2020.

Instituto Salvadoreño del Seguro Social (2021). Patronos y trabajadores reportados y cotizados al

ISSS. Retrieved from https://www.transparencia.gob.sv/institutions/isss/documents/estadisticas.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning. In Springer Texts in Statistics. Springer, New York.

Jung, J., Patnam, M and Ter-Martirosyan, A. (2018). An Algorithmic Crystal Ball: Forecasts-based on Machine Learning. IMF Working Paper.

Kuzin, V., Marcellino, M. and Schumacher, C. (2012). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area, International Journal of Forecasting Volume 27, Issue 2, April–June 2011, pp. 529–542.

Ministerio de Hacienda de El Salvador. (2021). Resultados Fiscales 2020 y Proyecciones de Mediano y Largo Plazo 2021–2030.

Mitchell, T. (2010). Artificial Neural Networks [PowerPoint presentation], Machine Learning 10-701, Machine Learning Department, Carnegie Mellon University. Retrieved from http://www.cs.cmu.edu/~epxing/Class/10701-10s/Lecture/lecture7.pdf.

Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial, Neurorobot 7: 21. Published online December 4, 2013. doi: 10.3389/fnbot.2013.00021.

Organisation for Economic Co-operation and Development, Production of Total Industry in the United Kingdom [GBRPROINDMISMEI]. Retrieved June 20, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GBRPROINDMISMEI.

Schumacher, C. (2014). MIDAS and bridge equations, Bundesbank Discussion Paper No 26/2014.

Sick, B. (2018). Advanced Studies in Applied Statistics (WBL) [PowerPoint presentation], ETHZ Applied Multivariate Statistics. Retrieved from https://ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material-1719/Multivariate/slides12_boosting_presented.pdf.

Statistical Institute of Belize (SIB) (2021). Statistics. Retrieved January 2021, from http://sib.org.bz/statistics/.

Tiffin, A. (2016). Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon. IMF Working Paper.

U.S. Bureau of Labor Statistics, Unemployment Rate - Hispanic or Latino [LNS14000009]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS14000009.

U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE]. Retrieved June 21, 2021, from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE.

UC Business Analytics (2018). Gradient Boosting Machines. Retrieved from http://uc-r.github.io/gbm_regression.

World Bank (2021). Overall level of statistical capacity. Retrieved January 2021, from https://data.worldbank.org/indicator.

Annexes

Annex 1. Variables Used for the Belize Nowcasting Exercise

Variable	Description	Classification
GDP_TOT_A	GDP, seasonally adjusted (constant 2000 prices – BZ$ million)	Flow
GDP_AG	GDP: agriculture, hunting and forestry (constant 2000 prices – BZ$ million)	Flow
GDP_FI	GDP: fishing (constant 2000 prices – BZ$ million)	Flow
GDP_MA	GDP: manufacturing (incl. mining and quarrying) (constant 2000 prices – BZ$ million)	Flow
GDP_EL	GDP: electricity and water (constant 2000 prices – BZ$ million)	Flow
GDP_CO	GDP: construction (constant 2000 prices – BZ$ million)	Flow
GDP_WH	GDP: wholesale and retail trade; repair (constant 2000 prices – BZ$ million)	Flow
GDP_HR	GDP: hotels and restaurants (constant 2000 prices – BZ$ million)	Flow
GDP_TR	GDP: transport and communication (constant 2000 prices – BZ$ million)	Flow
GDP_OT	GDP: other private services exc. FISIM 1) (constant 2000 prices – BZ$ million)	Flow
GDP_PO	GDP: producers of government services (constant 2000 prices – BZ$ million)	Flow
GDP_AL	GDP: all industries at basic prices (constant 2000 prices – BZ$ million)	Flow
GDP_TA	GDP: taxes on products (constant 2000 prices – BZ$ million)	Flow
CPI_AL	CPI: all items (index)	Index
CPI_FO	CPI: food and non-alcoholic beverages (index)	Index
CPI_AL	CPI: alcoholic beverages and tobacco (index)	Index
CPI_CL	CPI: clothing and footwear (index)	Index
CPI_HO	CPI: housing, water, electricity, gas, and other fuels (index)	Index
CPI_FU	CPI: furnishing, household equipment and routine household maintenance (index)	Index
CPI_HE	CPI: health (index)	Index
CPI_TR	CPI: transport (index)	Index
CPI_CO	CPI: communication (index)	Index
CPI_RE	CPI: recreation and culture (index)	Index

Variable	Description	Classification
CPI_ED	CPI: education (index)	Index
CPI_RE	CPI: restaurants and hotels (index)	Index
CPI_MI	CPI: misc. goods and services (index)	Index
EXP_FO	EXPORTS: food and live animals (BZE $ million)	Flow
EXP_BE	EXPORTS: beverages and tobacco (BZE $ million)	Flow
EXP_CR	EXPORTS: crude materials (BZE $ million)	Flow
EXP_MI	EXPORTS: mineral fuels and lub. (BZE $ million)	Flow
EXP_OI	EXPORTS: oils and fats (BZE $ million)	Flow
EXP_CH	EXPORTS: chemical products (BZE $ million)	Flow
EXP_MA	EXPORTS: manufactured goods (BZE $ million)	Flow
EXP_MAC	EXPORTS: mach. and transp. eqt. (BZE $ million)	Flow
EXP_OM	EXPORTS: other manufactures (BZE $ million)	Flow
EXP_CO	EXPORTS: commodities n.e.s (BZE $ million)	Flow
EXP_COM	EXPORTS: commercial processing zone (BZE $ million)	Flow
EXP_EP	EXPORTS: export processing zone (BZE $ million)	Flow
EXP_PE	EXPORTS: personal goods (BZE $ million)	Flow
EXP_TR	EXPORTS: transshipment (BZE $ million)	Flow
IMP_FO	IMPORTS: food and live animals (BZE $ million)	Flow
IMP_BE	IMPORTS: beverages and tobacco (BZE $ million)	Flow
IMP_CR	IMPORTS: crude materials (BZE $ million)	Flow
IMP_MI	IMPORTS: mineral fuels and lub. (BZE $ million)	Flow
IMP_OI	IMPORTS: oils and fats (BZE $ million)	Flow
IMP_CH	IMPORTS: chemical products (BZE $ million)	Flow
IMP_MA	IMPORTS: manufactured goods (BZE $ million)	Flow
IMP_MAC	IMPORTS: mach. and transp. eqt. (BZE $ million)	Flow
IMP_OM	IMPORTS: other manufactures (BZE $ million)	Flow
IMP_CO	IMPORTS: commodities n.e.s (BZE $ million)	Flow
IMP_COM	IMPORTS: commercial processing zone (BZE $ million)	Flow
IMP_EP	IMPORTS: export processing zone (BZE $ million)	Flow
IMP_PE	IMPORTS: personal goods (BZE $ million)	Flow
IMP_TR	IMPORTS: transshipment (BZE $ million)	Flow
MON_CG	Monetary: central government net ($'000)	stock
MON_NDC	Monetary: net domestic credit ($'000)	stock
MON_M1	Monetary: money supply (M1) ($'000)	stock

Variable	Description	Classification
MON_QU	Monetary: quasi-money ($'000)	stock
MON_M2	Monetary: money supply (M2) ($'000)	stock
MON_DEP_D	Monetary: deposit rates-demand	Rate
MON_DEP_SC	Monetary: deposit rates-savings/checking	Rate
MON_DEP_S	Monetary: deposit rates-savings	Rate
MON_DEP_T	Monetary: deposit rates-time	Rate
MON_DEP_WA	Monetary: deposit rates-weighted average	Rate
MON_LEN_P	Monetary: lending rates-personal	Rate
MON_LEN_C	Monetary: lending rates-commercial	Rate
MON_LEN_RC	Monetary: lending rates-residential construction	Rate
MON_LEN_O	Monetary: lending rates-other	Rate
MON_LEN_WA	Monetary: lending rates-weighted average	Rate
G_EXP_CU	Expenditure: current (BZE $ million)	Flow
G_EXP_CA	Expenditure: capital (BZE $ million)	Flow
G_OVE	Balances: overall (BZE $ million)	Flow
UNEM_US	US unemployment rate, percent, monthly, seasonally adjusted	Rate
UNEM_US_LA	US unemployment rate - Hispanic or Latino, percent, monthly, seasonally adjusted	Rate
EFFR_US	Effective federal funds rate, percent, monthly, not seasonally adjusted	Rate
IPI_US	Industrial production index, index 2012=100, monthly, seasonally adjusted	Index
IPM_US	Industrial production: manufacturing (NAICS), index 2012=100, monthly, seasonally adjusted	Index
MTB_6	6-month Treasury Bill: secondary market rate, percent, monthly, not seasonally adjusted	Rate
G_REV_TAX	Revenue: tax revenue ($'000)	Flow
G_REV_INP	Revenue: income and profits ($'000)	Flow
G_REV_CGR	Revenue: central government revenue	Flow
G_REV_TGS	Revenue: taxes on goods and services ($'000)	Flow
G_REV_ITT	Revenue: international trade and transactions ($'000)	Flow
SUP_SD	Sugar production: sugarcane deliveries (long tons)	Flow
SUP_SP	Sugar production: sugar production (long tons)	Flow
SUP_MP	Sugar production: molasses production (long tons)	Flow
CIP_CD	Citrus production: citrus deliveries (boxes)	Flow

Variable	Description	Classification
CIP_CJ	Citrus production: citrus juices ('000 ps)	Flow
ODP_BA	Other domestic products: banana (metric tons)	Flow
ODP_MP	Other domestic products: marine products ('000 lbs)	Flow
ODP_GA	Other domestic products: garments ('000 lbs)	Flow
ODP_PB	Other domestic products: petroleum (barrels)	Flow
TOU_A	Tourist arrivals: air	Flow
TOU_LAN	Tourist arrivals: land	Flow
TOU_SEA	Tourist arrivals: sea	Flow
TOU_SOV	Tourist arrivals: stay-over visitors	Flow
TOU_CSD	Tourist arrivals: cruise ship disembarkations	Flow
UK_IP	Production of total industry in the United Kingdom, index 2015=100, monthly, seasonally adjusted	Index
MON_AG	Monetary: loans and advances to the agricultural sector: sugar ($'000)	Stock
MON_CI	Monetary: loans and advances to the agricultural sector: citrus ($'000)	Stock
MON_GR	Monetary: loans and advances to the agricultural sector: grains ($'000)	Stock
MON_BAN	Monetary: loans and advances to the agricultural sector: bananas ($'000)	Stock
MON_CDI	Monetary: loans and advances to the agricultural sector: cattle and dairy ($'000)	Stock
MON_POE	Monetary: loans and advances to the agricultural sector: poultry and eggs ($'000)	Stock
MON_PA	Monetary: loans and advances to the agricultural sector: papayas ($'000)	Stock
MON_OTH	Monetary: loans and advances to the agricultural sector: other ($'000)	Stock
BOP_REM	BOP-secondary income: credit (mostly remittances)	Flow
GDP_TOT_A_1	Lag of GDP, seasonally adjusted (constant 2000 prices – BZ$ million)	Flow
GDP_AG_1	Lag of GDP: agriculture, hunting and forestry (constant 2000 prices – BZ$ million)	Flow
GDP_FI_1	Lag of GDP: fishing (constant 2000 prices – BZ$ million)	Flow
GDP_MA_1	Lag of GDP: manufacturing (incl. mining and quarrying) (constant 2000 prices – BZ$ million)	Flow
GDP_EL_1	Lag of GDP: electricity and water (constant 2000 prices – BZ$ million)	Flow
GDP_CO_1	Lag of GDP: construction (constant 2000 prices – BZ$ million)	Flow

Variable	Description	Classification
GDP_WH_1	Lag of GDP: wholesale and retail trade; repair (constant 2000 prices – BZ$ million)	Flow
GDP_HR_1	Lag of GDP: hotels and restaurants (constant 2000 prices – BZ$ million)	Flow
GDP_TR_1	Lag of GDP: transport and communication (constant 2000 prices – BZ$ million)	Flow
GDP_OT_1	Lag of GDP: other private services exc. FISIM 1) (constant 2000 prices – BZ$ million)	Flow
GDP_PO_1	Lag of GDP: producers of government services (constant 2000 prices – BZ$ million)	Flow
GDP_AL_1	Lag of GDP: all industries at basic prices (constant 2000 prices – BZ$ million)	Flow
GDP_TA_1	Lag of GDP: taxes on products (constant 2000 prices – BZ$ million)	Flow

Source: Prepared by the authors on the basis of CBB, SIB, BTB and FED-FRED.

Annex 2. Variables Used for the El Salvador Nowcasting Exercise

Variable	Description	Classification
GDP_D	GDP-total	Index
CE_TOT_VVV	FUSADES - business confidence - all sectors - sales volume	Index
CE_TOT_VI	FUSADES - business confidence - all sectors - investment volume	Index
CE_IND_VVV	FUSADES - business confidence - industry - sales volume	Index
CE_IND_VI	FUSADES - business confidence - industry - investment volume	Index
CE_CON_AG	FUSADES - business confidence - construction - overall activity	Index
CE_CON_VI	FUSADES - business confidence - construction – investment volume	Index
CE_COM_VVV	FUSADES - business confidence - commerce - sales volume	Index
CE_COM_VI	FUSADES - business confidence - commerce - investment volume	Index
CE_SER_VVV	FUSADES - business confidence - services - sales volume	Index
CE_SER_VI	FUSADES - business confidence - services - investment volume	Index
PI_D	BCR: industrial production index. Seasonally adjusted series	Index
IVAE_TOT_D	Total IVAE	Index
IVAE_AG_D	IVAE: agriculture, livestock, forestry and fisheries	Index
IVAE_IN_D	IVAE: manufacturing industries, mining and quarrying, and other industrial activities	Index
IVAE_CO_D	IVAE: construction	Index
IVAE_CT_D	IVAE: commerce, transportation and storage, lodging and food service activities	Index
IVAE_IC_D	IVAE: information and communications	Index

Variable	Description	Classification
IVAE_AF_D	IVAE: financial and insurance activities	Index
IVAE_AI_D	IVAE: real estate activities	Index
IVAE_AP_D	IVAE: professional, scientific, technical, administrative, support and other services activities	Index
IVAE_AA_D	IVAE: public administration and defense, education, health and social assistance activities	Index
PRO_ENER	Electricity production (thousands of kWh)	Flow
CON_ENER	Electricity consumption (thousands of kWh)	Flow
CON_APA_CEM	Apparent cement consumption (thousands of 42.5 kg bags)	Flow
TRAN_CAR_AE	Air freight (thousands of kg)	Flow
TRAN_CAR_MR	Maritime freight (thousands of metric tons)	Flow
ENT_PAS	Passenger arrivals (number of persons)	Flow
SAL_PAS	Passenger departures (number of persons)	Flow
EXP_CO	Exports: green coffee and other unprocessed coffees	Flow
EXP_AG	Exports: agriculture, livestock, forestry and fishing - remaining products.	Flow
EXP_MC	Exports: mining and quarrying	Flow
EXP_AC	Exports: cane sugar and other sugars	Flow
EXP_CP	Exports: processed coffee	Flow
EXP_IM	Exports: manufacturing - remaining products	Flow
EXP_SE	Exports: electricity, gas, steam, and air conditioning supply	Flow
EXP_CM	Exports: wholesale and retail trade, repair of motor vehicles and motorcycles	Flow
EXP_MO	Exports: maquila - other products	Flow
EXP_MP	Exports: maquila - knitwear	Flow
EXP_MT	Exports: maquila - textile products	Flow
IMP_D_BN	Detailed imports: consumer non-durables	Flow
IMP_D_BD	Detailed imports: consumer durables	Flow
IMP_D_IM	Detailed imports: intermediate goods, manufacturing industry	Flow
IMP_D_BA	Detailed imports: intermediate goods, agriculture and livestock	Flow
IMP_D_BC	Detailed imports: intermediate goods, construction	Flow
IMP_D_BO	Detailed imports: intermediate goods, other	Flow
IMP_D_BM	Detailed imports: capital goods, manufacturing industry	Flow
IMP_D_BT	Detailed imports: capital goods, transp. and communication	Flow
IMP_D_BCA	Detailed imports: capital goods, agriculture and livestock	Flow

Variable	Description	Classification
IMP_D_BCC	Detailed imports: capital goods, construction	Flow
IMP_D_BIC	Detailed imports: capital goods, commerce	Flow
IMP_D_BCS	Detailed imports: capital goods, services	Flow
IMP_D_BE	Detailed imports: capital goods, electricity, water and services	Flow
IMP_D_BCB	Detailed imports: capital goods, banking	Flow
IMP_D_BCO	Detailed imports: capital goods, other	Flow
IMP_D_MA	Detailed imports: maquila	Flow
REM	Remittances	Flow
INGT_ISR	NFPS tax revenues: income tax (net)	Flow
INGT_PAT	NFPS tax revenues: property	Flow
INGT_TRP	NFPS tax revenues: property transfers	Flow
INGT_IMP	NFPS tax revenues: imports	Flow
INGT_COP	NFPS tax revenues: consumption of goods	Flow
INGT_USE	NFPS tax revenues: use of services	Flow
INGT_TFS	NFPS tax revenues: stamp duties	Flow
INGT_IVA	NFPS tax revenues: value added tax (VAT) (net)	Flow
INGT_OTR	NFPS tax revenues: other	Flow
INGT_FOV	NFPS tax revenues: special contributions (FOVIAL)	Flow
INGC_CSS	NFPS current revenues: social security contributions	Flow
INGC_NOT	NFPS current revenues: non-tax	Flow
INGC_SEM	NFPS current revenues: public companies' operating surplus	Flow
INGC_TRP	NFPS current revenues: transfers from public financial institutions	Flow
GPC	NFPS public consumption expenditure	Flow
GIP	NFPS public investment	Flow
BFG	NFPS overall fiscal balance (excluding grants)	Flow
M3	M3	Stock
CTC	Consumer credit: credit cards	Stock
CCO	Consumer credit: other	Stock
CCV	Housing credit	Stock
CAG	Credit for agriculture, livestock, forestry, hunting and fishing	Stock
CCON	Construction credit	Stock
CPI	Credit for industry	Stock
CCR	Credit for commerce, restaurants and hotels	Stock
CTA	Credit for transportation, warehousing and communications	Stock

Variable	Description	Classification
COS	Credit for other services	Stock
COPN	Credit: other unclassified loans	Stock
TIP_30	30-day passive interest rate	Rate
TIP_180	180-day deposit interest rate	Rate
TPR1	One-year lending rate	Rate
TPR1_MAS	Lending rate above one year	Rate
ISSS_PRIM	ISSS contributors, primary sector	Stock
ISSS_SEC	ISSS contributors, secondary sector	Stock
ISSS_TER	ISSS contributors, tertiary sector	Stock
UNEM_US	US unemployment rate, percent, monthly, seasonally adjusted	Rate
UNEM_US_LA	US unemployment rate - Hispanic or Latino, percent, monthly, seasonally adjusted	Rate
EFFR_US	Effective federal funds rate, percent, monthly, not seasonally adjusted	Rate
IPI_US	Industrial production index, index 2012=100, monthly, seasonally adjusted	Index
IPM_US	Industrial production: manufacturing (NAICS), index 2012=100, monthly, seasonally adjusted	Index
MTB_6	6-month Treasury Bill: secondary market rate, percent, monthly, not seasonally adjusted	Rate

Source: Prepared by the authors on the basis of BCR, ISSS, FUSADES and FED-FRED.

Annex 3. Variables selected by the lasso regression for Belize

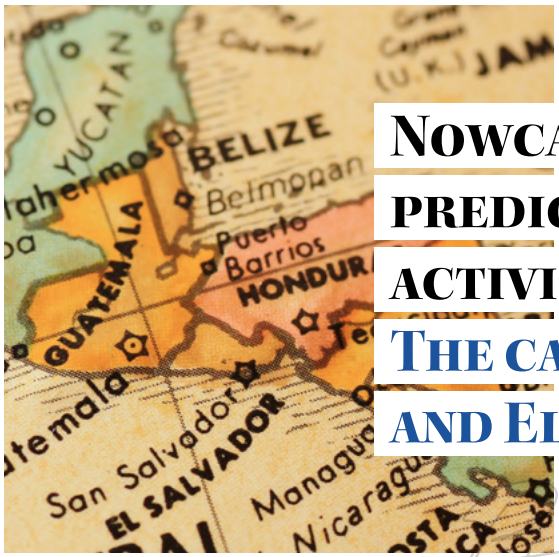
Variable	Description
CPI_HO	Consumer price index: housing, water, electricity, gas, and other fuels
CPI_MI	Consumer price index: misc. goods and services
EXP_FO	Exports: food and live animals (BZE $ million)
IMP_FO	Imports: food and live animals (BZE $ million)
IMP_CR	Imports: crude materials (BZE $ million)
IMP_MI	Imports: mineral fuels and lub. (BZE $ million)
IMP_OI	Imports: oils and fats (BZE $ million)
IMP_MA	Imports: manufactured goods (BZE $ million)
IMP_COM	Imports: commercial processing zone (BZE $ million)
IMP_PE	Imports: personal goods (BZE $ million)
MON_LEN_RC	Monetary: lending rates - residential construction
G_OVE	Government balances: overall (BZE $ million)
UNEM_US_LA	US unemployment rate - Hispanic or Latino, percent, monthly, seasonally adjusted
G_REV_TAX	Revenue: tax revenue ($'000)
G_REV_ITT	Revenue: international trade and transactions ($'000)
ODP_BA	Production: other domestic products: bananas (metric tons)
ODP_MP	Production: other domestic products: marine products ('000 lbs)
TOU_LAN	Tourist arrivals: land
TOU_SEA	Tourist arrivals: sea
TOU_SOV	Tourist arrivals: stay-over visitors
UK_IP	Production of total industry in the United Kingdom, index 2015=100, monthly, seasonally adjusted
MON_AG	Monetary: loans and advances to the agricultural sector: sugar ($'000)
MON_BAN	Monetary: loans and advances to the agricultural sector: bananas ($'000)
MON_OTH	Monetary: loans and advances to the agricultural sector: other ($'000)

Source: Prepared by the authors on the basis of BCR, ISSS, FUSADES and FED-FRED.

Annex 4. Variables Selected by the Lasso Regression for El Salvador

Variable	Description
CE_IND_VI	FUSADES - business confidence - industry – investment volume
CE_CON_AG	FUSADES - business confidence - construction - overall activity
IPI	BCR: industrial production index. Seasonally adjusted series
IVAE_TOT	IVAE: total IVAE
IVAE_CO	IVAE: construction
IVAE_CT	IVAE: commerce, transportation and storage, lodging and food service activities
IVAE_IC	IVAE: information and communications
IVAE_AF	IVAE: financial and insurance activities
IVAE_AI	IVAE: real estate activities
CON_ENER	Electricity consumption (thousands of kWh)
TRAN_CAR_MR	Maritime freight (thousands of metric tons)
EXP_AG	Exports: agriculture, livestock, forestry and fishing - remaining products
EXP_CM	Exports: wholesale and retail trade, repair of motor vehicles and motorcycles
EXP_MT	Exports: maquila, textile products
IMP_D_BA	Detailed imports: intermediate goods, agriculture and livestock
IMP_D_BCA	Detailed imports: capital goods, agriculture and livestock
IMP_D_BCC	Detailed imports: capital goods, construction
IMP_D_BCB	Detailed imports: capital goods, banking
INGT_ISR	NFPS tax revenues: income tax (net)
INGT_TRP	NFPS tax revenues: property transfers
INGT_COP	NFPS tax revenues: consumption of goods
INGT_IVA	NFPS tax revenues: value added tax (net)
INGT_OTR	NFPS tax revenues: other
INGC_CSS	NFPS current revenues: social security contributions
INGC_SEM	NFPS current revenues: public companies' operating surplus
GIP	NFPS public investment
CCO	Consumer credit: other
COS	Credit for other services
TPR1_MAS	One-year lending rate
ISSS_PRIM	Cotizantes ISSS sector primario

Source: Prepared by the authors on the basis of BCR, ISSS, FUSADES and FED-FRED.



Nowcasting to predict economic activity in real time: The cases of Belize and El Salvador

Belize and El Salvador

IDB
Inter-American Development Bank