A Machine-Learning Approach to Nowcast the GDP
in Sub-Saharan Africa

Draft (Please Do Not Circulate)

September 11, 2020

Karim Barhoumi, Seung Mo Choi, Tara Iyer, Jiakun Li, Franck Ouattara, Andrew Tiffin, Jiaxiong Yao[1]

Abstract

In order to aid policy decisions during the COVID-19 pandemic, we propose a machine learning approach to nowcast GDP growth in sub-Saharan Africa, a region where official statistics are released with considerable delays. We show that machine learning methods provide nowcasts with lower root mean square errors than standard benchmarks used in the literature. Nowcasts imply that the COVID-19 crisis initially had a large adverse impact on the region.

I.   INTRODUCTION

Effective policymaking, especially during the COVID-19 crisis, relies on timely assessments of the current state of the economy. The lack of timely information on key measures of economic activity, such as GDP, is particularly problematic during the COVID-19 crisis as the situation rapidly evolves. However, many countries in sub-Saharan Africa are developing economies with limited institutional capacity and do not necessarily compile and publish comprehensive and timely macroeconomic statistics. Official national accounts data are often released with considerable delays.

Nowcasting techniques can help to address the problem of data release delays in sub-Saharan Africa by providing a means of tracking current economic activity. Nowcasting models seek to estimate current and near future economic activity by extracting signals from indicators that are typically available at the same or higher frequencies that the outcome variable of interest. This allows for ongoing predictions of GDP growth, months or quarters before the official statistics would be released.

This study employs machine learning techniques to develop a nowcasting model to track quarterly GDP growth in sub-Saharan Africa. Machine learning techniques have gained attention as a rapidly expanding sub-field of statistics and are ideally suited to extracting more reliable signals from a large set of noisy high-frequency indicators. To be specific, machine learning techniques are an extension of non-parametric statistics and exploit historical (possibly

non-linear) statistical patterns to predict the output indicator (which is GDP in this case). While traditional econometric theory focuses more on issues of causality and accurately fitting models to the data, machine learning focuses more on producing more accurate predictions. Our framework is purely statistical, using historical relationships to estimate the implications of macroeconomic data as they are released. The algorithms are fit on a dataset of monthly and quarterly economic indicators in countries in sub-Saharan Africa to nowcast quarterly GDP growth.

The results indicate that machine learning methods are superior to standard benchmarks used in the literature. The region's real GDP growth is proxied by 11 countries whose quarterly GDP data are available from 2010:Q1, weighted by PPP GDPs. These countries are Angola, Botswana, Cameroon, Côte d'Ivoire, Ghana, Kenya, Lesotho, Namibia, Nigeria, South Africa, and Tanzania. These countries account for about ¾ of the region's PPP GDP. These countries account for around three-fourths of the region's PPP GDP. A suite of machine learning algorithms are estimated on the data, with the first 90 percent of the data used for training and tuning and the rest used to evaluate the performance of competing models including ordinary least squares (OLS). Using cross-validation, the chosen model is the one with the lowest root mean square error (RMSE) during this evaluation period. The findings indicate that the machine learning models perform better than the random walk, AR(1), and OLS models, and are able to closely track GDP growth in Sub-Saharan Africa.

The rest of the paper is structured as follows. Section II discusses the literature. Section III provides an overview of the machine learning methods and discusses the data used in the study. Section IV evaluations the performance of the framework. Section V interprets the results economically. Section VI concludes and outlines next steps in the research agenda.

II.   RELATED LITERATURE

The concept of nowcasting is relatively new to the economic literature and was developed around a decade ago. The seminal paper of Giannone et al. (2008) introduced the technique, and nowcasting models have subsequently been adopted at several Central Banks including the Federal Reserve. While the literature has focused primarily on nowcasting in advanced economies, some recent studies develop nowcasting models for developing economies including Turkey (Solmaz and Sanjani, 2015), Lebanon (Tiffin, 2016), India (Iyer and Gupta, 2019), and other developing economies (Marini, 2016; Narita and Yin, 2018). Using a dataset of high frequency indicators, this study contributes to the literature by being among the first to develop a nowcasting model to predict quarterly GDP growth in Sub-Saharan Africa.

Machine learning provides an alternative to the dynamic factor model methodology commonly used in the nowcasting literature. Factor-based models extract a small set of latent factors from a large set of indicators by exploiting the co-movement among variables (eg. Barhoumi et al , 2016, 2012, 2010, Giannone et al, 2008; Bok et al, 2017; Iyer and Gupta, 2019). While this approach has its merits, a drawback is that even though the factors are able to summarize the information available in the data, dynamic factor models are not designed to place more weight on variables that might individually be better predictors of the output indicator. The machine learning method, on the other hand, places more emphasis on variable

selection by employing algorithms to place greater weight on individual variables with more informative content.

III.   FRAMEWORK AND DATA

The machine learning models used in the study are the support vector machine (SVM), random forest, elastic net, stochastic gradient boosting trees, relevance vector machines (RVM), gaussian process, multivariate adaptive regression splines, and ensemble. The SVM is an algorithm that defines support vectors and employ a more robust (e-sensitive) error loss function. The random forest algorithm combines a large number of decision trees to approximate a non-linear non-additive regression. The elastic net model is a penalized regression model with coefficient shrinkage and feature elimination and trades some bias to improve out-of-sample performance. The ensemble is a meta-algorithm that ranks the other machine learning models and combines decisions from different machine learning techniques. The technique of cross-validation is used to evaluate the predictive ability of alternate nowcasting models estimated on the SSA data. The purpose of cross-validation is to gauge the out-of-sample performance of a machine learning model using only in-sample data. The appendix discusses details on the various machine learning models used in this study, as well as a discussion on cross-validation and data imputation methods.

To evaluate the performance of competing models, we use the root mean squared error (RMSE) criterion. The RMSE, which measures the distance between the actual time series and the predicted values, is commonly used in the literature to evaluate how close the nowcasts are to the data. The first 90 percent of data is used for tuning and training. The last 10 percent of data is used as a hold-out set to evaluate the performances of alternative models, including simple OLS. The model chosen is the one with the lowest root mean square error during the hold-out evaluation period. The key difference between cross-validation and holdout validation, which is often used in evaluating predictive ability, is that the former method takes advantage of the entire dataset by using all combinations of the testing and training sets. This produces an array of validation errors associated with that particular model, which then provides a gauge of its average out-of-sample performance. This metric can be used to help choose between different types of models.

We also analyze the Shapley decomposition, which are localized and help explain why the model generates particular predictions. Shapley values divide each model's prediction among the variables in a way that fairly represents their contributions across all possible subsets of variables, thereby highlighting the contribution of different variables to an individual prediction. For the machine learning model that we choose, Shapley values indicate which variables prompted the model to assign a higher probability for that variable (compared to the sample average) and will provide a quantitative guide as to each variable's relative contribution to the point prediction. The Shapley decomposition accounts for nonlinearities and attributes contributions to an individual nowcast. [We also depend on the Local Interpretable Model-agnostic Explanations (LIME) which linearizes the model around individual nowcasts and conducts sensitivity analysis to evaluate which features impact the prediction the most.]

There are 10 monthly predictors which are used to nowcast quarterly GDP growth. These include the Brent crude oil price, the FIBER industrial materials index, the business confidence index in South Africa, purchasing managers indices (PMIs) in South Africa, Nigeria, and China, new vehicles sold in South Africa, stock price indices in selected countries, and the REER (weighted average across countries).

IV.   EVALUATION OF THE FRAMEWORK

Figure 1 provides the year-on-year rolling quarterly real GDP growth, data and nowcasts based on the best machine learning model. All the machine learning algorithms including variants of the Elastic Net, Random Forest, and Support Vector Machine, are able to outperform standard benchmarks in the literature including the random walk, AR(1) processes, and OLS regressions.

Figure 1. Sub-Saharan Africa: Year-on-Year Quarterly Real GDP Growth, Data and Projections, Percent
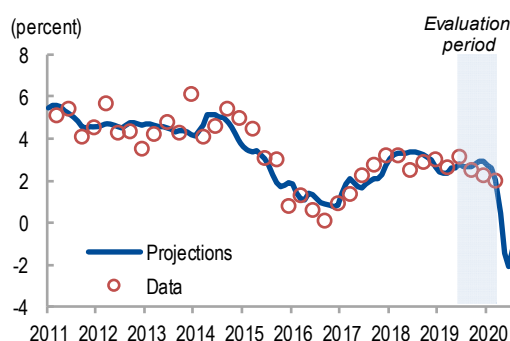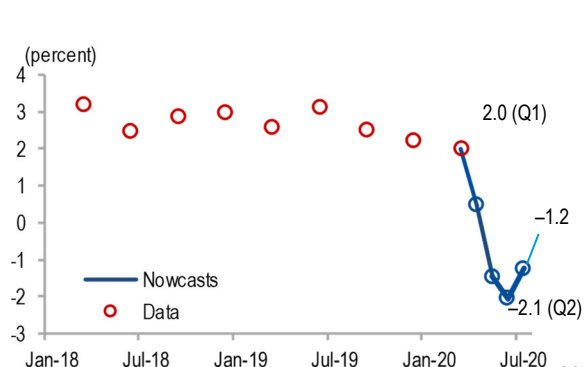

Sources: Haver; and IMF staff calculations.

V.   INTERPREATION OF THE NOWCASTS AND HIGH-FREQUENCY INDICATORS

Our nowcast for sub-Saharan-African growth in the quarter ending in July 2020 is -1.2 percent year-on-year (Figure 2). Monthly activity is projected to have rebounded in July after a decline in four consecutive months since March. The rebound reflects the fact that countries have continued to loosen containment measures, natural resource prices have continued to recover, and financial conditions have continued to improve. The Shapley Decomposition of projected year-on-year real GDP growth in the quarter ending in July is provided in Figure 3.

High-frequency macroeconomic indicators suggest that COVID-19 initially had an adverse impact in the region, although there have been signs of a gradual uptick in economic activity. In South Africa, sales of new vehicles fell by over 30 percent year-on-year in March 2020 and business confidence has substantially deteriorated. South Africa's purchasing managers' index (PMI) dropped by 30 percent year-on-year to around 18 in April 2020 before recovering to around 41 in July 2020 (Figure 4). Similarly, Nigeria's PMI declined to around 37 in April 2020, before recovering to 50 in July 2020. Ghana and Kenya's PMIs underwent similar falls and gradual upticks (Figure 5). Nitrogen dioxide, a proxy for economic activity, and a form of toxic emissions by vehicles, burning fossil fuels, and other industrial activities, has also markedly decreased in the region (Figure 6).

Figure 2. Sub-Saharan Africa: Year-on-Year Rolling Quarterly Real GDP Growth, Data and Nowcasts

Sources: Haver; IMF internal databases; and IMF staff calculations.

Figure 3. Sub-Saharan Africa: Shapley Decomposition of Projected Year-on-Year Real GDP Growth in the Quarter Ending in July
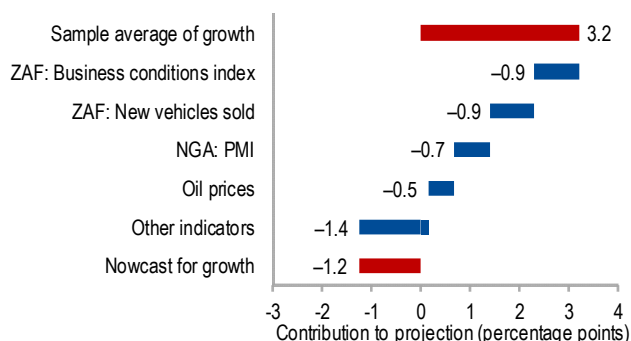

Note: The Shapley decomposition breaks the projection into contributions from predictors by considering nonlinearities and the joint impact of interrelated variables.

Sources: Haver; IMF internal databases; and IMF staff calculations.

The fall in oil prices has exacerbated the economic slowdown in the region, while exports have increased over the past few months. Crude oil production has decreased in several countries over the past few months (Figure 7). Oil prices, which tend to move together with Sub-Saharan African average growth, are recovering from the April bottom but remain lower than the pre-Covid-19 level (Figure 8). Non-oil commodity prices also remain lower than the pre-Covid-19 levels (Figure 9). While not used in nowcasting framework, seaborne exports from sub-Saharan Africa have recovered strongly since early May, after a year-on-year drop by 20 percent, while currencies have weakened (Figures 10 and 11). However, seaborne imports declined by almost 10 percent, year-on-year, in late July.

There have been pressures on budgets across the region and monetary policy has become accommodative. The average planned COVID-19-related fiscal spending in sub-Saharan African countries is lower than in advanced economies or emerging markets globally (Figure 12). Confronting the COVID-19 shock, sub-Saharan African countries allocated additional fiscal resources to the health sector (0.8 percent of GDP) as well as the non-health sector (1.8 percent of GDP). Furthermore, fiscal revenues across several countries have declined over the course of the pandemic. For instance, South Africa's fiscal revenue in June 2020 declined by almost 40 percent from the same month the last year (Figure 13). Several sub-Saharan African countries have loosened their monetary policy stance since March, as indicated by the decline of policy interest rates since March in countries including Ghana, Madagascar, and South Africa (Figure 14). The pressure on food prices has been limited although food inflation is rising in countries including Angola and Ethiopia (Figure 15).

Figure 4. South Africa and Nigeria: Purchasing Managers Index (Expansion if >50; Contraction if <50)
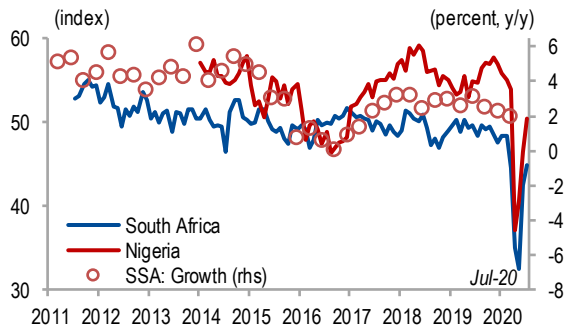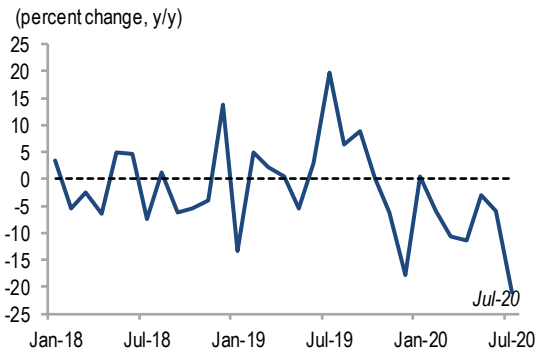

Sources: Haver; and IMF staff calculations.

Figure 5. Ghana and Kenya: Purchasing Managers Index (Expansion if >50; Contraction if <50)
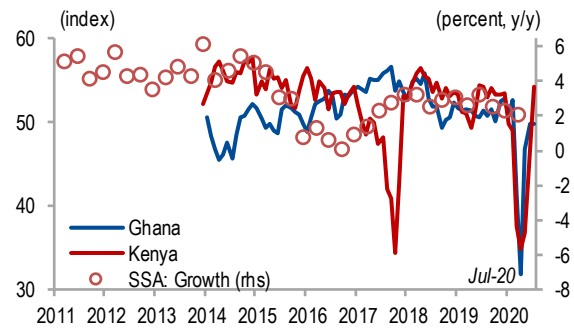
Sources: Haver; and IMF staff calculations.

Figure 6. Sub-Saharan Africa: Nitrogen Dioxide (NO$_2$) Levels

Note: Figure shows the (simple) average across 45 sub-Saharan African countries. NO$_2$ is a form of toxic emissions by vehicles, burning fossil fuels, and other industrial activities.

Sources: National Aeronautics and Space Administration (NASA); and IMF staff calculations.

Figure 7. Sub-Saharan Africa: Crude Oil Production
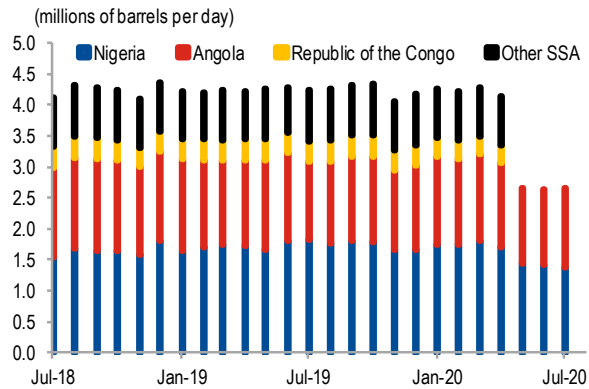
Note: "Other SSA" includes Cameroon, Chad, Côte d'Ivoire, Equatorial Guinea, Gabon, and Ghana. Data for Angola and Nigeria are available up to June. Data for other countries are available up to March.
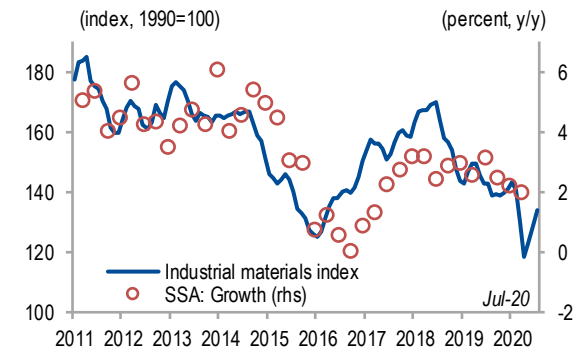
Source: Trading Economics

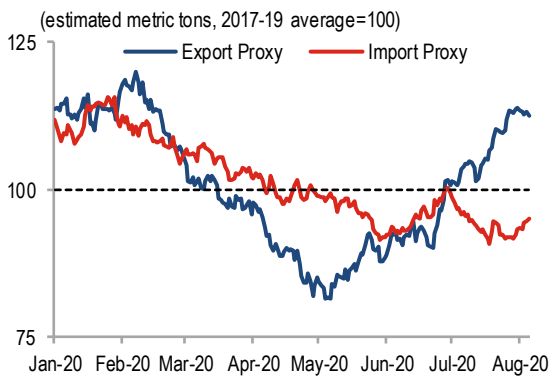**Figure 8. WTI Crude Oil Spot Market Price**
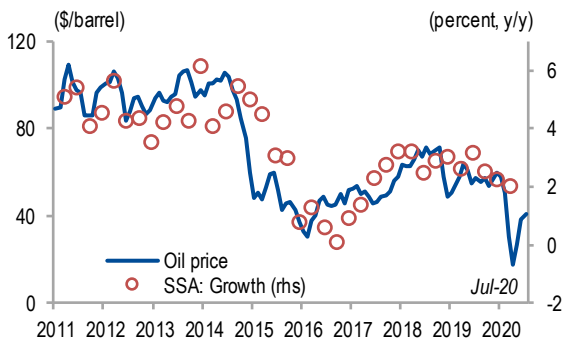
Sources: Haver; and IMF staff calculations.

Figure 9. FIBER Industrial Materials Index, All Items
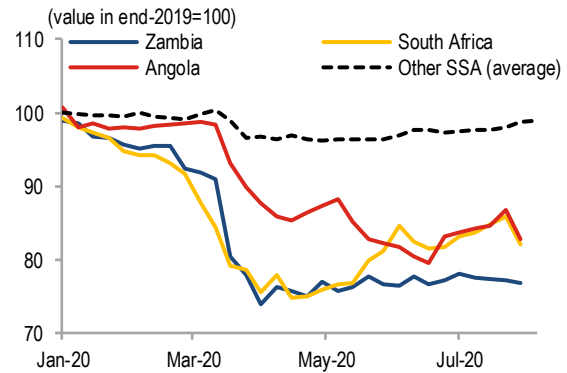

Sources: Haver; and IMF staff calculations.

Figure 10. Africa: Metric Tons of Seaborne Exports and Imports, Estimated Based on Automatic Identification System (AIS)

(estimated metric tons, 2017-19 average=100)

Note: Data are constructed from the radio signals that the global vessel fleet emits for navigational safety purposes.

Sources: Cerdeiro, Komaromi, Liu and Saeed (2020); AIS data collected by MarineTraffic.
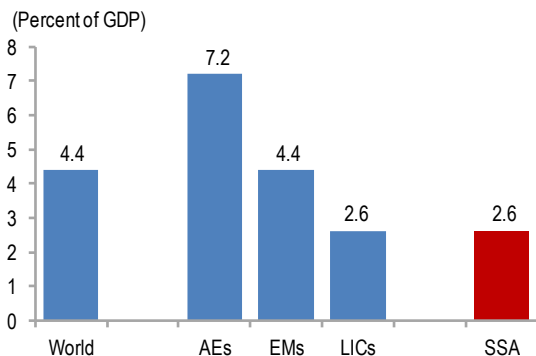
Figure 11. Selected Sub-Saharan African Economies: Currency Value against US Dollar

(value in end-2019=100)

Note: "Other SSA" includes Côte d'Ivoire, Gabon, Ghana, Kenya, Mauritius, Nigeria, Rwanda, Senegal, Tanzania, and Uganda.
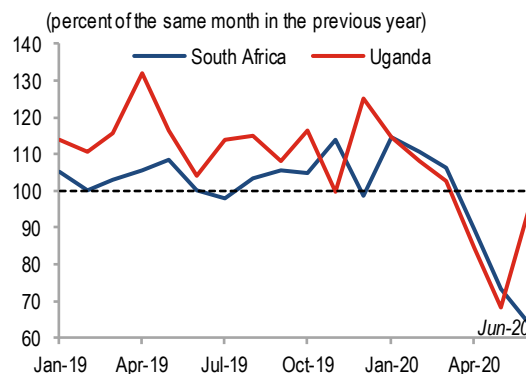
Sources: Haver; and IMF staff calculations

Figure 12. Selected Economies: Additional Fiscal Resources, as of mid-July

(Percent of GDP)

Note: Figure shows (simple) averages. Figure includes above-the-line or on-budget measures in response to COVID-19 as additional spending or foregone revenue.

Source: IMF internal database, "Policy Responses to COVID-19 and Related Shocks," Fiscal section, Question 1.1.1.2.

Figure 13. South Africa and Uganda: Fiscal Revenue

(percent of the same month in the previous year)

Sources: Haver; and IMF staff calculations.

On the other hand, Uganda's fiscal revenue in June has recovered close to the last year's level.
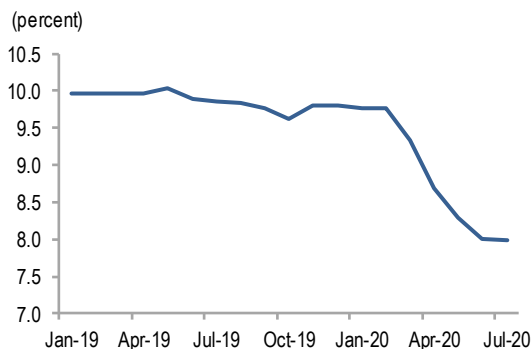
Figure 14. Selected Sub-Saharan African Economies: Food Inflation
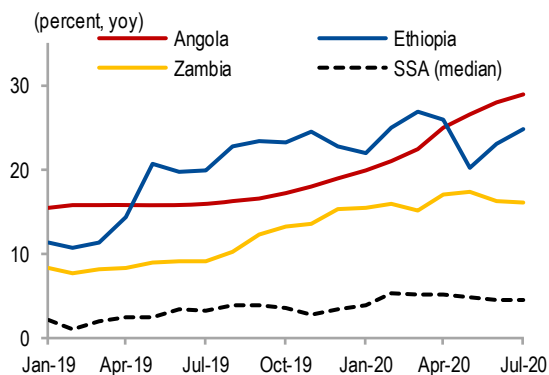
Note: The SSA median is for 17 countries whose monthly food prices are available from January 2019 to July 2020.

Sources: Haver; and IMF staff calculations

Figure 15. Selected Sub-Saharan African Countries: Policy Interest Rates, Average

Note: Figure shows the (simple) average across sub-Saharan African countries classified as having de-facto floating exchange rates. These countries are Ghana, Madagascar, Mauritius, Mozambique, South Africa, Uganda, and Zambia. Seychelles is excluded due to data issues.

Source: Trading Economics; and IMF staff calculations.

VI.   CONCLUSION

This study sought to nowcast GDP growth in Sub-Saharan Africa using machine learning methods. The findings from this study indicate that machine learning methods, including elastic net, support vector machine, and random forest, are able to produce superior nowcasts compared to traditional regression methods. In context of the COVID-19 pandemic, it is hoped that the models used in this study can provide policymakers with a means to track the current state of economic activity as the situation rapidly evolves. A next step in the research agenda is to assess the informational content of more financial indicators as well as employ data on google search trends and from satellite images in tracking economic activity in Africa.

REFERENCES

Barhoumi, K., Ferrara, L., (2016). A World Trade Leading Index (WTLI), Economic Letters, vol.146, pp.111-115.

Barhoumi, K., Ferrara, L., Darné, O., (2012). Nowcasting the German GDP: A comparison of bridge and factor models, Journal of Policy Modeling.

Barhoumi, K., Ferrara, L., Darné, O, (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? . Journal of Forecasting.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A., Tambalotti, A. (2017). Macroeconomic Nowcasting and Forecasting with Big Data. Federal Reserve Bank of New York Staff Reports, Staff Report No. 830.

Giannone, D., Reichlin, L., Small, D. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. Journal of Monetary Economics, 55, 665-676.

Iyer, T. and Gupta, A. (2019). Nowcasting Economic Growth in India: The Role of Rainfall. Asian Development Bank Economics Working Paper Series No. 593.

Marini, M.(2016). Nowcasting Annual National Accounts with Quarterly Indicators: An Assessment of Widely Used Benchmarking Methods. IMF Working Paper No. 16/71.

Narita, F. and Yin, R. (2018). In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries. IMF Working Papers 18/286.

Solmaz, S. and Taheri Sanjani, M. (2015). How External Factors Affect Domestic Economy: Nowcasting an Emerging Market. IMF Working Paper No. 15/269.

Tiffin, A. (2016). Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon. IMF Working Paper No. 16/56.