United Nations
Economic Commission for Africa

Nowcasting GDP using Big Data and Machine Learning

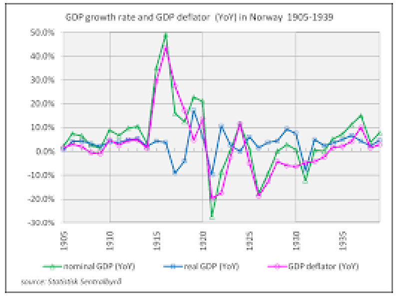
The African Experience

Issoufou Seidou Sanda

African Centre for Statistics
United Nations Economic Commission for Africa

StatsTalk – October 31, 2023

ideas for a prosperous Africa
IdeastoAction



Nowcasting: The estimation or prediction of the current or near-term state of a target variable

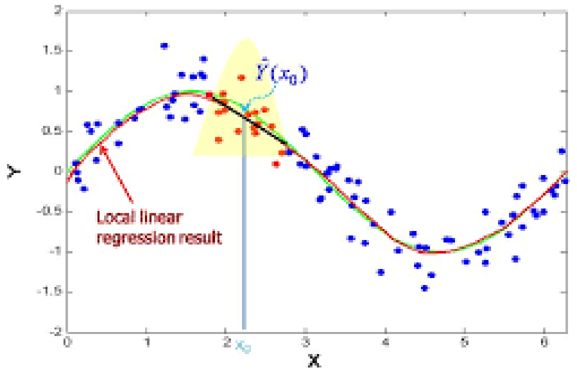
Nowcasting is a contraction of the terms "now" and "forecasting." It refers to the estimation or prediction of the current or near-term state of a target variable using available real-time data and predictive analytics.

- ✓ Nowcasting provides timely and up-to-date estimates of the current state of the economy.
- ✓ Official GDP figures are often released with a significant delay: no real-time information about the economy's performance.

Nowcasting GDP: Traditional methods

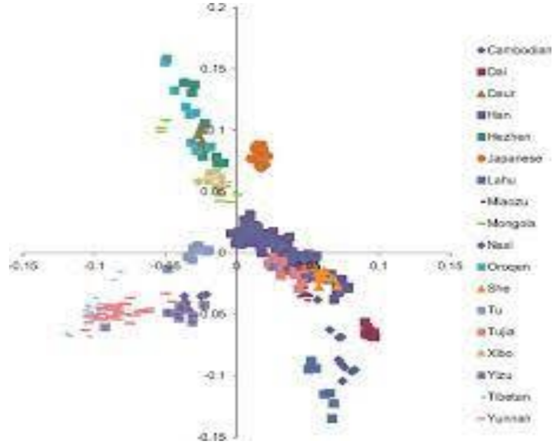
1. Statistical Models:

1. **Linear Regression Models:** These models use past GDP data and other related indicators to make short-term predictions.
2. **Vector AutoRegressive (VAR) Models:** These capture linear interdependencies among multiple time series. They can be extended into Vector Error Correction Models (VECM) if cointegration exists among variables.



2. Factor Models:

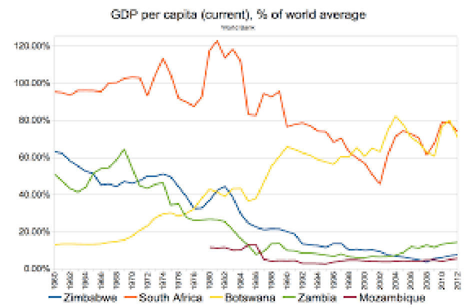
1. **Dynamic Factor Models (DFM):** They extract common trends from large datasets, using them to make predictions.
2. **Principal Component Analysis (PCA):** A method to reduce the dimensionality of the data, capturing the main variations in the dataset with fewer variables.



Nowcasting GDP: Traditional methods

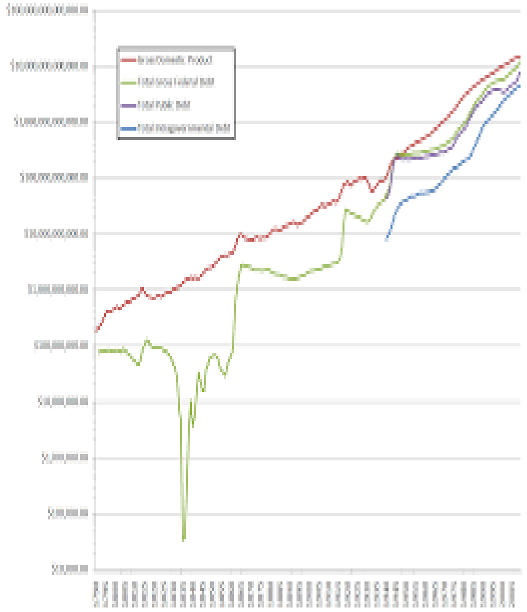
3. Bridge Models:

These models relate quarterly GDP growth to monthly indicators. Monthly indicators can provide information about the current quarter's GDP growth before the official release.



4. Professional Forecasts:

Institutions like banks, research bodies, and international organizations often make their GDP predictions. These predictions can be aggregated to provide a consensus forecast, which can be useful for nowcasting.



5. Judgmental Adjustments:

Expert economists can adjust model-based forecasts based on qualitative information or recent events not captured by the model.

Nowcasting GDP: Traditional methods

5. Moving Averages and Exponential Smoothing:

These methods capture trends in data over time, smoothing out short-term fluctuations to give a clearer view of the underlying trend.

6. Benchmarking and Temporal Disaggregation:

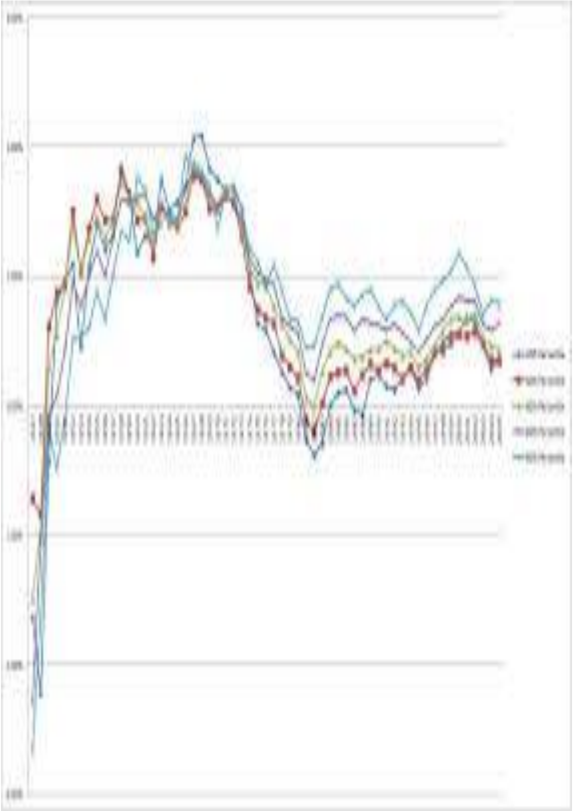
Sometimes, data at a desired frequency (e.g., monthly) might be estimated using data at another frequency (e.g., yearly). Techniques like the Chow-Lin procedure can be used to disaggregate the data.

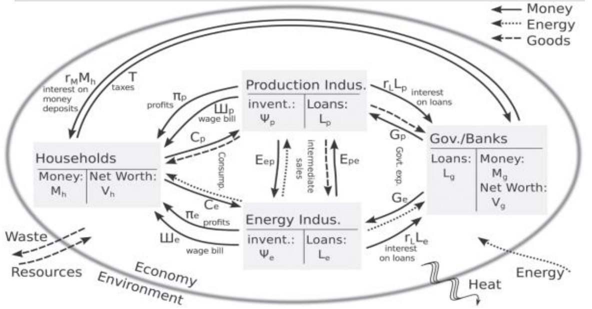
7. Surveys and Soft Indicators:

1. Surveys about business confidence, consumer confidence, and purchasing managers' indexes (PMI) provide insights into economic activity that can be useful for nowcasting.

8. Economic Theories and Models:

1. Structural or theoretical models based on economic theories (like the IS-LM model) can be used to make short-term predictions.



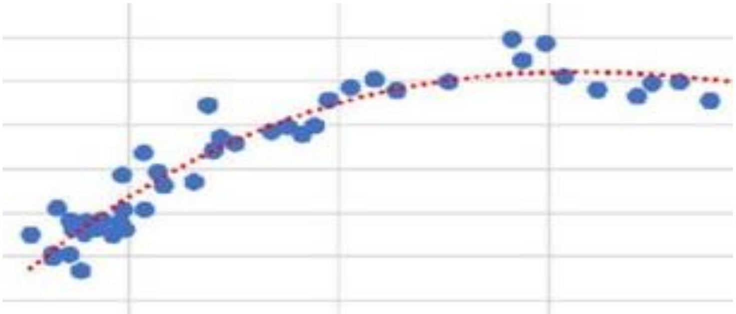


Importance of Accurate GDP Predictions for African Economies

- **Economic Policy and Planning:** Accurate GDP predictions enable governments to formulate timely and effective fiscal and monetary policies, ensuring economic stability and growth.

- **Foreign Investment and Aid Decisions:** Reliable GDP forecasts can influence foreign investment inflows and aid allocation, as investors and international organizations rely on these figures to gauge economic health and potential return on investment.

- **Debt Management:** Proper GDP projections assist in prudent debt management, ensuring that countries do not over-borrow and can sustainably service their debts.



Accurate GDP prediction is crucial for the Sustainable Development Goals (SDGs) because it provides a measure of a country's economic health and potential for growth. By forecasting GDP, policymakers can:

- allocate resources effectively
- prioritize investments
- implement strategies to achieve the SDGs.

A precise GDP forecast helps in assessing progress towards eradicating poverty, ensuring quality education, and promoting sustainable industries.

Challenges in GDP Predictions for African Economies

- **Data Scarcity:** Many African nations face challenges in consistent data collection and reporting, leading to gaps that can skew GDP predictions.

- **Informal Economies:** A significant portion of economic activity in Africa occurs outside the formal sector, making it challenging to measure and include in GDP calculations accurately.

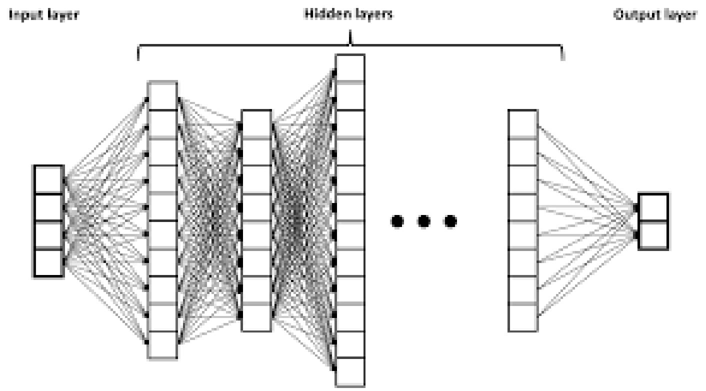
- **Rapid Changes:** African economies can be subject to swift socio-political and economic shifts, which can introduce volatility and unpredictability into GDP forecasts.

Why Big Data and Machine Learning are Revolutionizing GDP Nowcasting

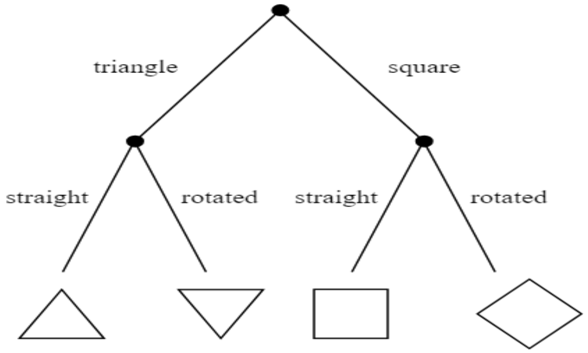
- **Vast Data Processing:** Big data tools can handle and process enormous datasets in real-time, capturing minute details and variances that traditional models might miss, leading to more accurate nowcasts.
- **Adaptive Learning:** Machine learning algorithms continuously learn and adapt from new data, improving prediction accuracy over time without requiring manual intervention or recalibration.
- **Incorporation of Diverse Data Sources:** Big data allows for the inclusion of non-traditional and high-frequency data sources, such as social media sentiment, satellite imagery, and online transactions, providing a broader and more immediate view of economic activity.
- **Complex Pattern Recognition:** Machine learning can detect intricate patterns and relationships in vast datasets, which might be imperceptible through traditional statistical methods, enhancing the precision of nowcasting efforts.

Main Models and Methods Using Big Data and Machine Learning for GDP Nowcasting

- **Neural Networks:**
  - Deep learning models that can capture non-linear relationships in large datasets.
  - Include architectures like feedforward, recurrent neural networks (RNN), and long short-term memory (LSTM) networks.

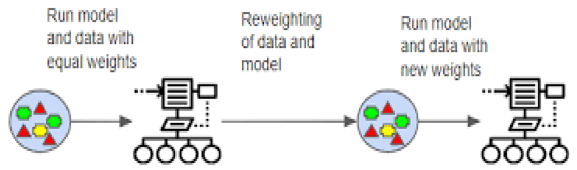


- **Random Forests:**
  - Ensemble learning method that uses multiple decision trees.
  - Excellent for handling high-dimensional data and determining feature importance.

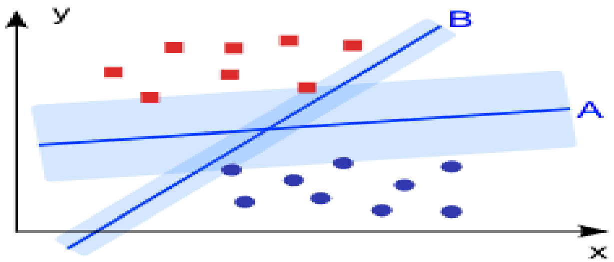


Main Models and Methods Using Big Data and Machine Learning for GDP Nowcasting

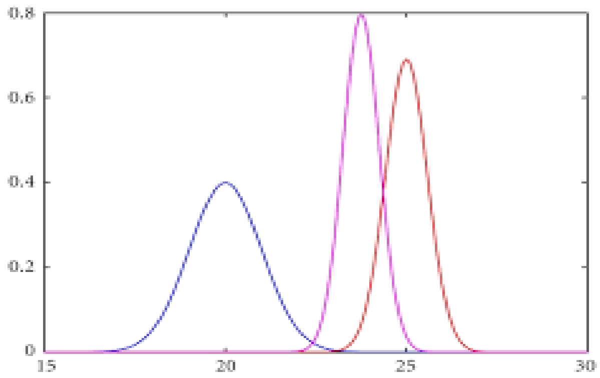
- **Gradient Boosting Machines (GBM):**
  - Sequentially builds weak prediction models, improving accuracy by reducing bias and variance.



- **Support Vector Machines (SVM):**
  - Used for regression and classification tasks, excels in high-dimensional spaces.

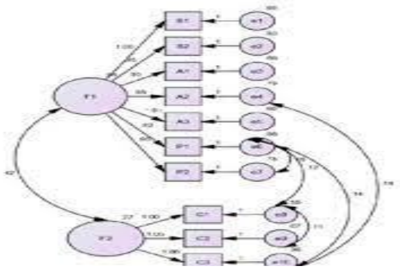


- **Bayesian Models:**
  - Incorporate prior knowledge and uncertainty in the predictions.
  - Examples include Bayesian Neural Networks and Gaussian Processes.

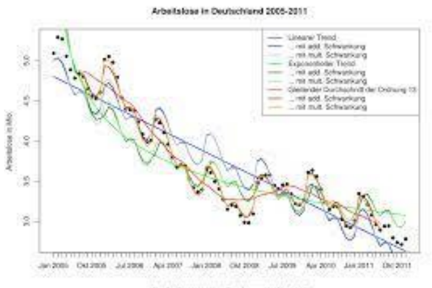


Main Models and Methods Using Big Data and Machine Learning for GDP Nowcasting

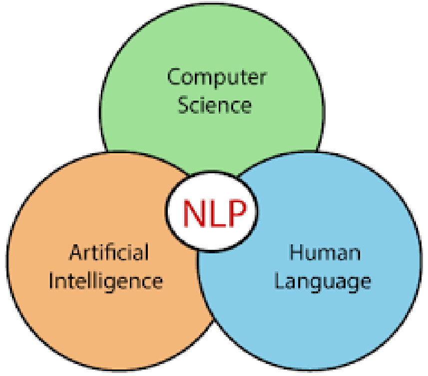
- **Factor Models with Machine Learning:**
  - Extensions of traditional factor models using machine learning techniques to capture latent factors from big data.



- **Time Series Models with ML Enhancements:**
  - ARIMA, VAR, and State-Space models integrated with machine learning components for improved nowcasting.

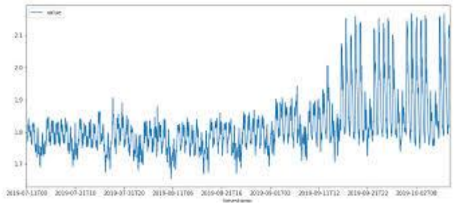


- **Natural Language Processing (NLP):**
  - Analyzing textual data, such as news articles or social media sentiments, which can provide insights into economic conditions.

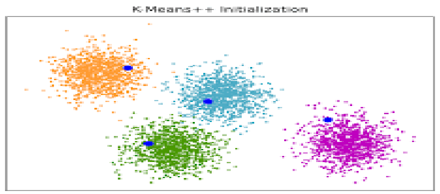


Main Models and Methods Using Big Data and Machine Learning for GDP Nowcasting

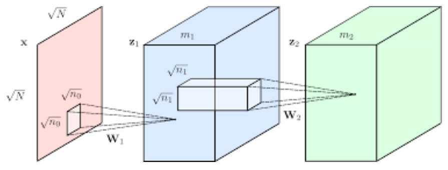
- **Anomaly Detection:**
  - Identifying outliers or unusual patterns in economic data which might indicate significant economic events or errors.



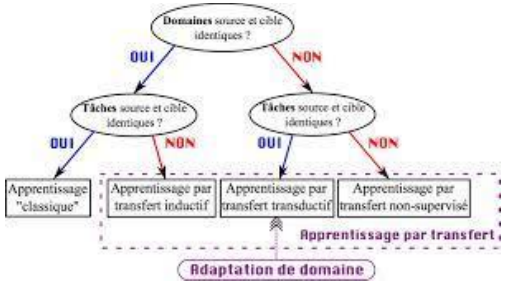
- **K-means Clustering:**
  - Segmenting data into clusters to identify patterns or trends in various economic segments.



- **Convolutional Neural Networks (CNN):**
  - Primarily used for image data, can be applied to satellite imagery to assess economic activities, such as infrastructure development or agricultural yield predictions.

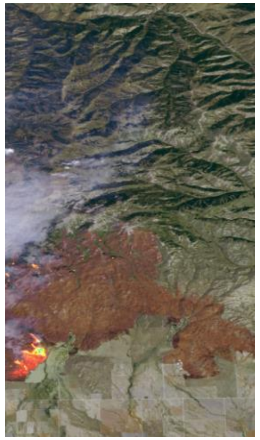


- **Transfer Learning:**
  - Applying pre-trained models on related tasks to accelerate the training process and potentially improve accuracy



Main Data Sources Used for Nowcasting GDP with Big Data and Machine Learning

- **Satellite Imagery:**
  - Monitoring infrastructure development, urban growth, and agricultural yields.
- **Social Media Data:**
  - Gauging public sentiment, consumer confidence, and tracking economic discussions or trends.
- **E-Commerce and Online Transactions:**
  - Real-time insights into consumer behavior, retail sales, and business activity.
- **Internet Search Data:**
  - Using search engine queries to predict economic trends based on user interest and concerns.



Main Data Sources Used for Nowcasting GDP with Big Data and Machine Learning

- **Electronic Payments and Point-of-Sale (POS) Data:**
  - Monitoring daily transactions to infer consumer spending patterns and trends.
- **High-Frequency Trading Data:**
  - Reflecting financial market sentiments and potential leading indicators for economic activity.
- **Mobile Phone Data:**
  - Tracking mobility patterns, communication frequencies, and even mobile money transactions in some regions.
- **Utilities Consumption:**
  - Electricity, water, and gas usage patterns can indicate industrial and residential economic activity.

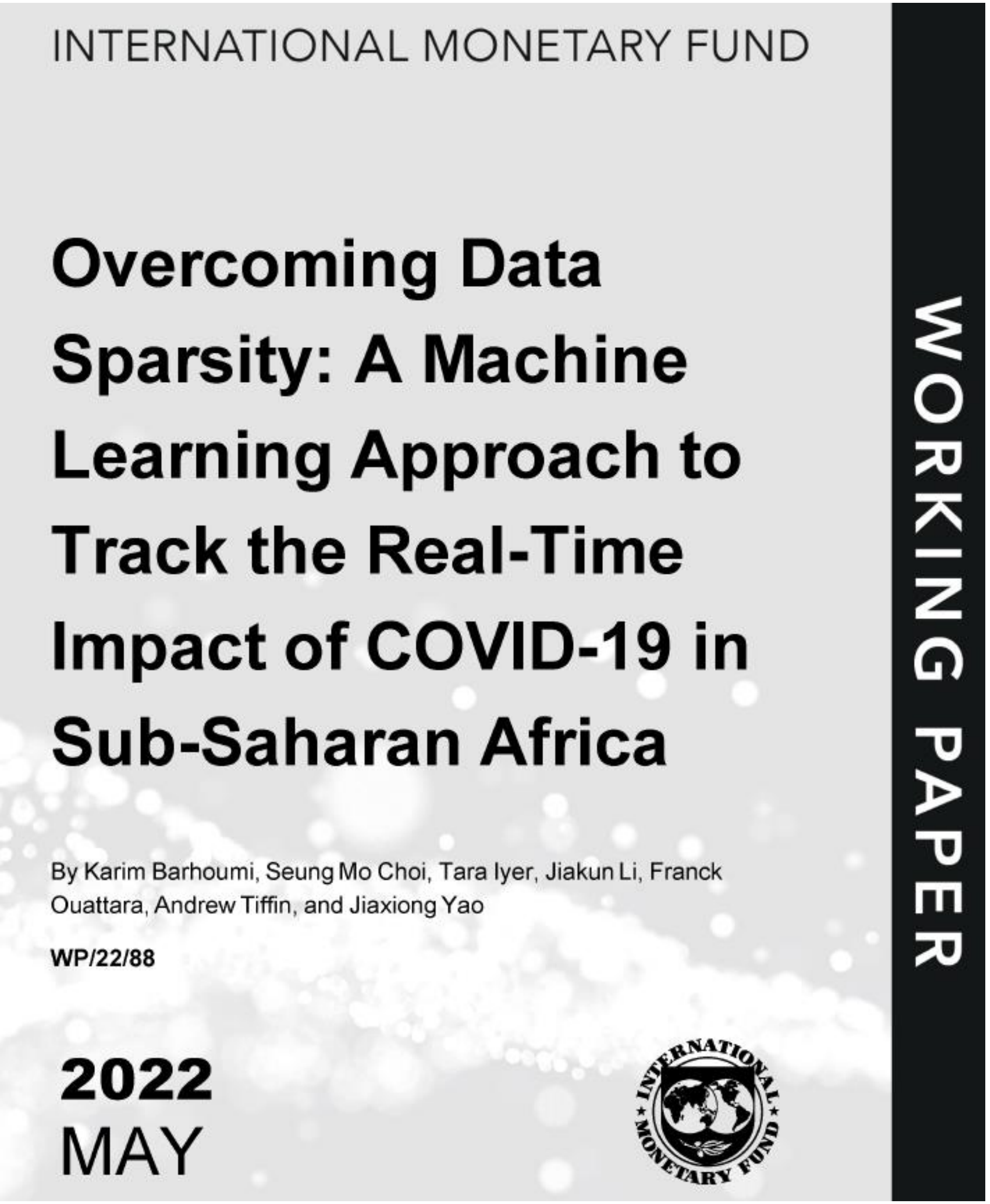
Main Data Sources Used for Nowcasting GDP with Big Data and Machine Learning

- **Sensors and IoT Devices:**
  - Real-time monitoring of supply chains, transportation, manufacturing, and more.
- **Customs and Trade Data:**
  - International shipping and customs data can provide insights into import/export trends.
- **Job Listings and Online Employment Portals:**
  - Indicators of labor market health, skills demand, and potential economic growth.
- **Financial Apps and Fintech Data:**
  - Insights into personal finance behaviors, savings, loans, and investment patterns.

Main Data Sources Used for Nowcasting GDP with Big Data and Machine Learning

- **Real Estate Platforms and Listings:**
  - Tracking housing market trends, construction activities, and property investments.

- **Government and Institutional Databases:**
  - High-frequency public data releases, like weekly unemployment claims or sector-specific reports.
- **Surveys and Polls:**
  - Regular surveys capturing business sentiment, consumer confidence, or sector-specific insights.


INTERNATIONAL MONETARY FUND

WORKING PAPER

Overcoming Data Sparsity: A Machine Learning Approach to Track the Real-Time Impact of COVID-19 in Sub-Saharan Africa

By Karim Barhoumi, Seung Mo Choi, Tara Iyer, Jiakun Li, Franck Ouattara, Andrew Tiffin, and Jiaxiong Yao

WP/22/88

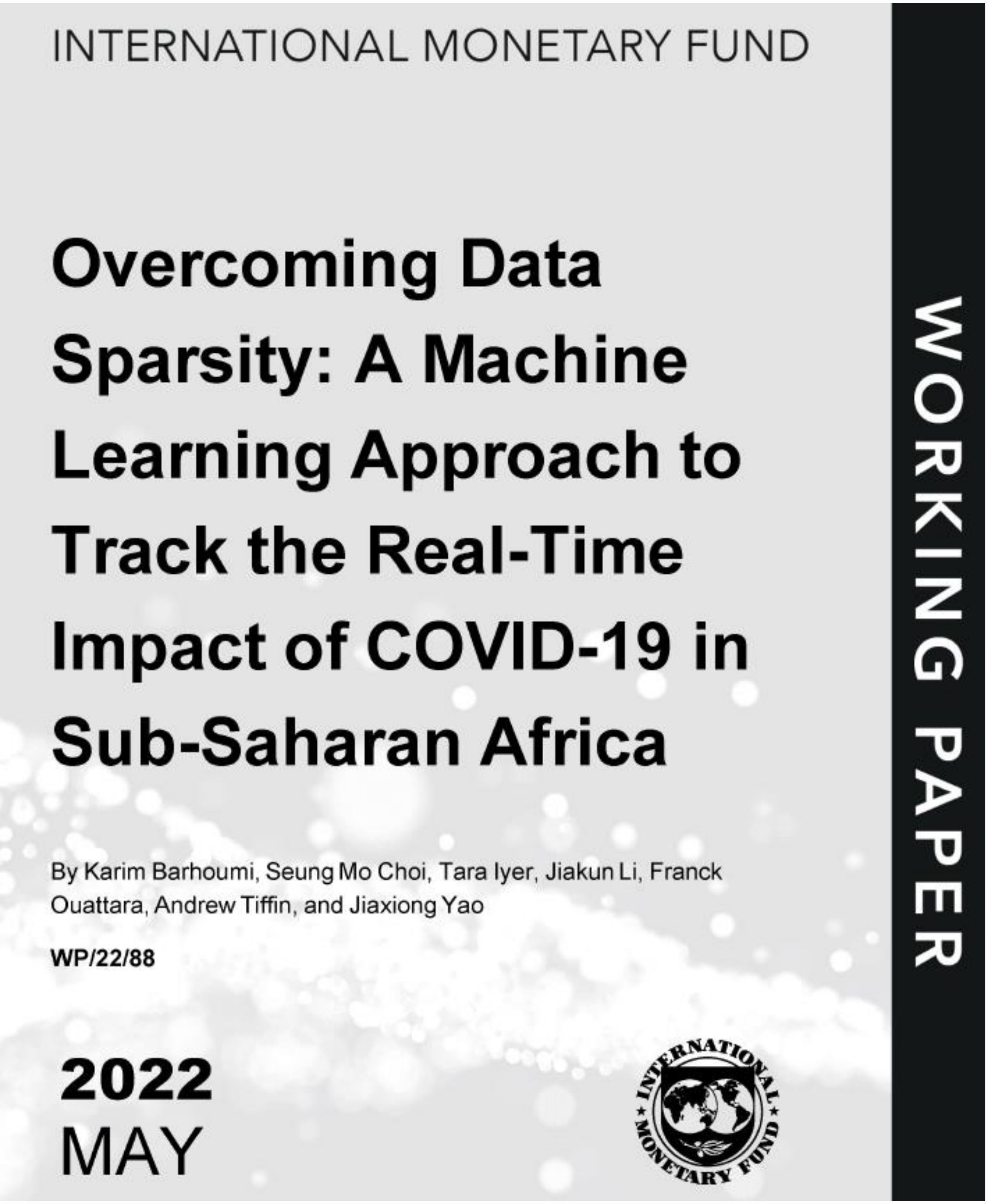
2022
MAY

The document is a comprehensive exploration of using machine learning to understand the economic impact of COVID-19 in sub-Saharan Africa, especially in the context of data sparsity and delays in official statistics.

The paper discusses machine learning techniques, such as gradient boosting, which are often seen as a "black box" due to their complexity. The authors leverage "interpretable machine learning" methods, including Shapley decompositions, to understand the framework's nowcast projections better.

Shapley values, based on cooperative game theory, are used to estimate the contribution of each predictor to an individual projection. These values indicate which variables caused the model to deviate from the sample average, providing a quantitative guide for each variable's relative contribution to the prediction.

The paper provides illustrative examples, with Nigeria being a primary focus. The predictors for Nigeria include its real effective exchange rate and its imports from the US and India.


INTERNATIONAL MONETARY FUND

Overcoming Data Sparsity: A Machine Learning Approach to Track the Real-Time Impact of COVID-19 in Sub-Saharan Africa

By Karim Barhoumi, Seung Mo Choi, Tara Iyer, Jiakun Li, Franck Ouattara, Andrew Tiffin, and Jiaxiong Yao

WP/22/88

2022
MAY

WORKING PAPER

Selecting Predictors

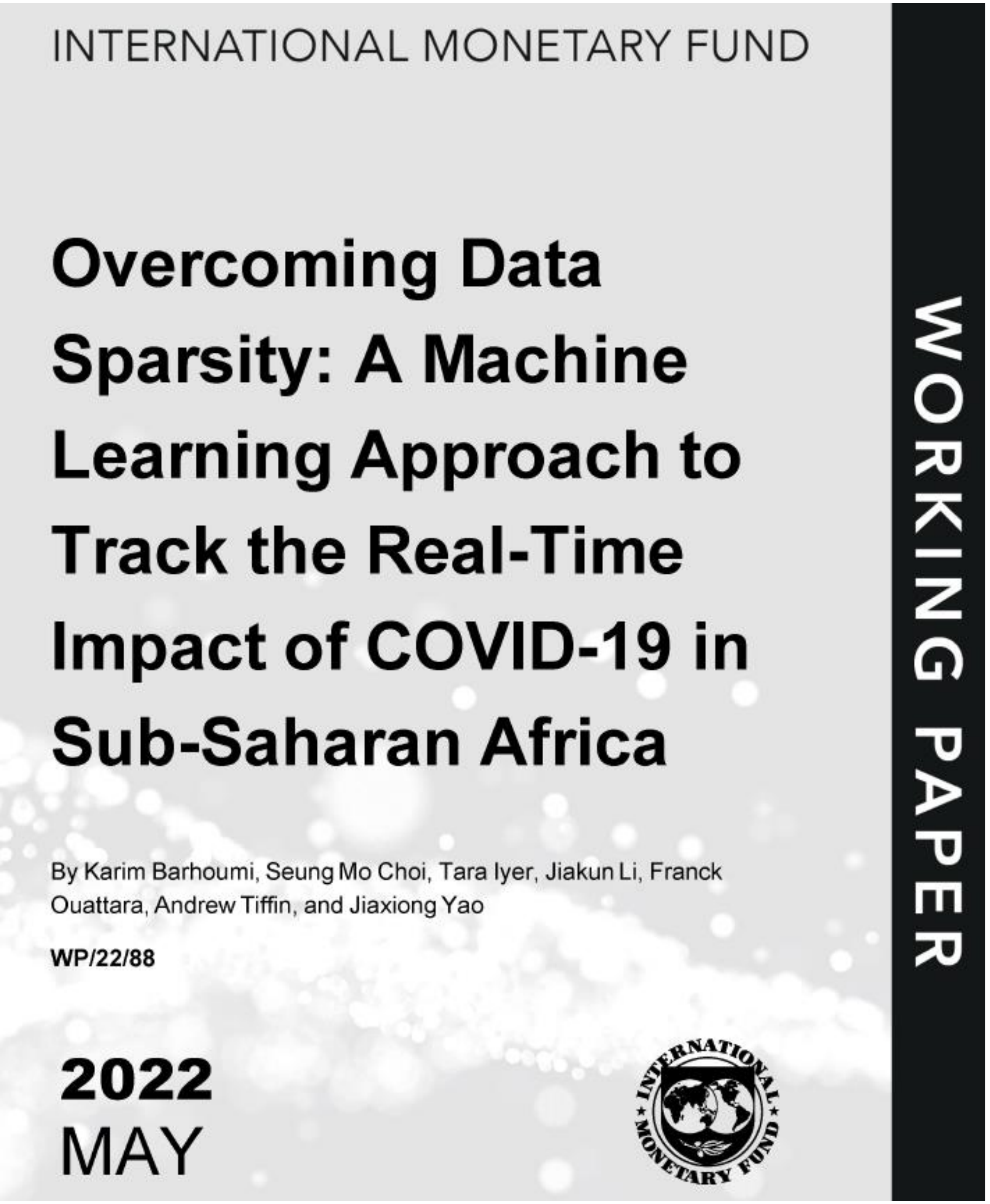
- ✓ Predictors are chosen based on their historical relationship to GDP growth, whether linear or non-linear.

- ✓ These predictors should be released in a timely manner, ideally well before the official GDP figures are published.

- ✓ The data for these predictors should be available for a sufficiently long period, ideally matching the length of the available GDP series. For instance, the global oil price is considered a good predictor for Nigeria's GDP growth because it moves in sync with it, is available almost in real-time

Selecting the Best Model ("Horseracing")

- ✓ The nowcasting framework evaluates the performance of over 30 different types of models on a holdout test set. Each model is first "tuned" on a training set.
- ✓ There's also an ensemble option, which is a second-layer algorithm that combines the predictions of the previous models.
- ✓ The model with the lowest Root Mean Square Error (RMSE) during the hold-out evaluation period is typically selected as the best model.


INTERNATIONAL MONETARY FUND

Overcoming Data Sparsity: A Machine Learning Approach to Track the Real-Time Impact of COVID-19 in Sub-Saharan Africa

By Karim Barhoumi, Seung Mo Choi, Tara Iyer, Jiakun Li, Franck Ouattara, Andrew Tiffin, and Jiaxiong Yao

WP/22/88

2022 MAY

WORKING PAPER

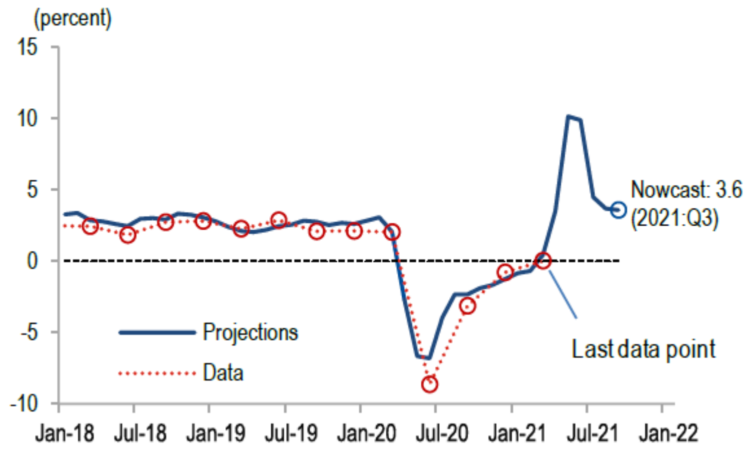
Nowcasting:

Once the best model is identified, it is re-estimated using the entire sample, not just the test set.

The predictions from this re-estimated model serve as the basis for the nowcast.

Figure 9: Sub-Saharan Africa: Year-on-year Rolling Quarterly Real GDP Growth, Data and Projections

After an unprecedented contraction in the second quarter of 2020 at the height of the pandemic, sub-Saharan Africa's economy continued to recover.



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

Source: Barhoumi, K., Choi, S. M., Iyer, T., Li, J., Ouattara, F., Tiffin, A., ... & Yao, J. (2022). Overcoming Data Sparsity: A Machine Learning Approach to Track the Real-Time Impact of COVID-19 in Sub-Saharan Africa (No. 2022-2088). International Monetary Fund.

South African Reserve Bank Working Paper Series WP/21/01

Nowcasting South African GDP using a suite of statistical models

Byron Botha, Geordie Reid, Tim Olds, Daan Steenkamp and Rossouw van Jaarsveld

Authorised for distribution by Witness Simbanegavi

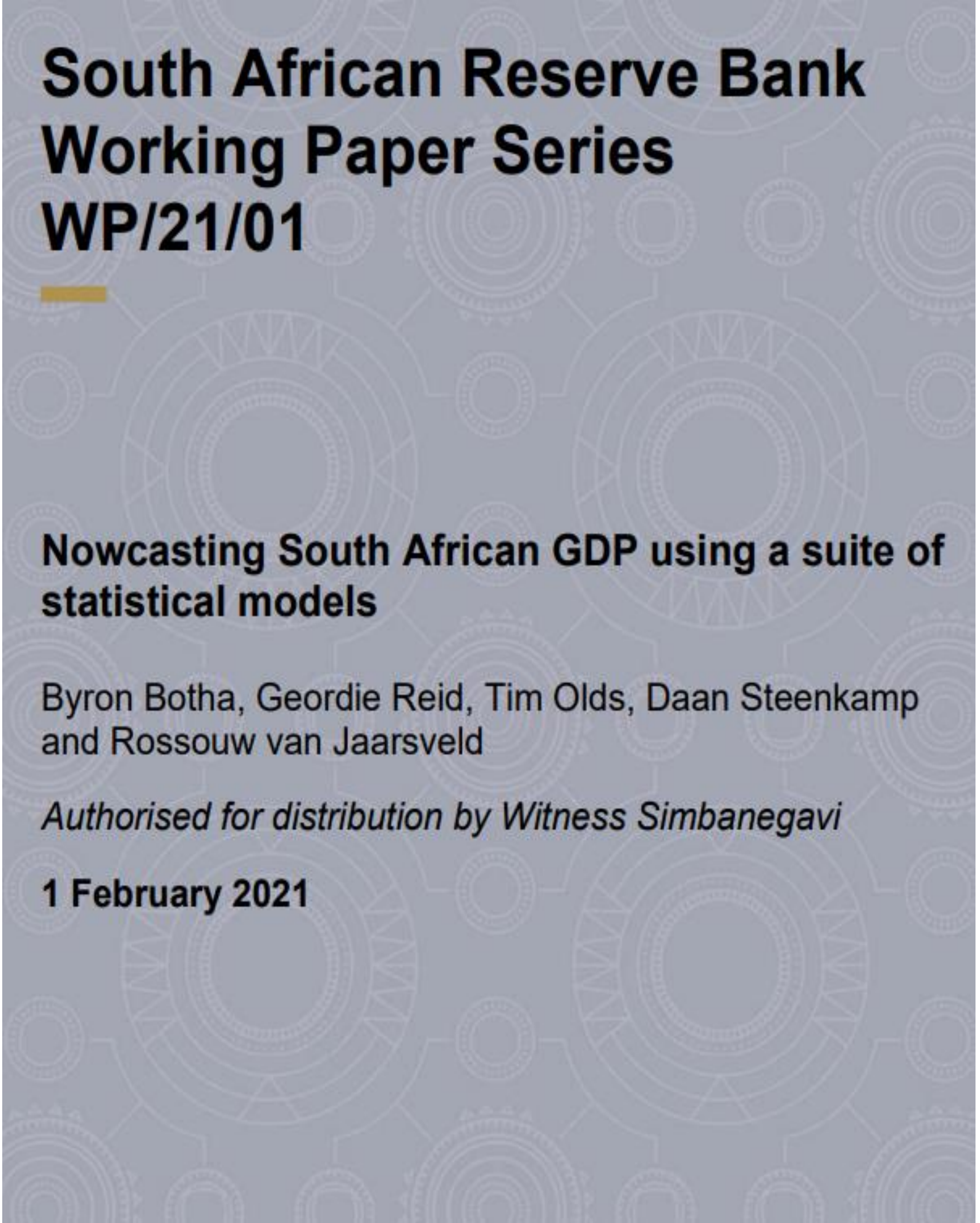
1 February 2021

Methods used:

1. Mixed Data Sampling (MIDAS): MIDAS (Mixed Data Sampling) is one of the techniques mentioned. It has become increasingly popular for nowcasting quarterly GDP. MIDAS forecasts quarterly data using monthly predictors by applying a function of weights to the monthly data at different lags.

2. Lasso Model: The Lasso model is a type of regression analysis method that performs both variable selection and regularization. The document provides mathematical formulations for the quarterly Lasso estimate and the MIDAS Lasso estimate. The Lasso model is defined in a way that it penalizes the absolute size of the regression coefficients, ensuring that only the most important predictors are included in the model.

3. Variables Used in Machine Learning Models: The document lists several variables used in the machine learning models for nowcasting GDP. These include Gross domestic product at market prices, Long-term government bond yield, Trade Activity Index, ABSA PMI, JP Morgan global manufacturing PMI, Real effective exchange rate of the rand, CPI, Value of building plans passed, Job advertisement space, Leading business cycle indicator, Industrial production, New Passenger Vehicle Sales, Electricity consumed in South Africa, Real Retail Trade Sales, M3, Physical volume of manufacturing production, and many more



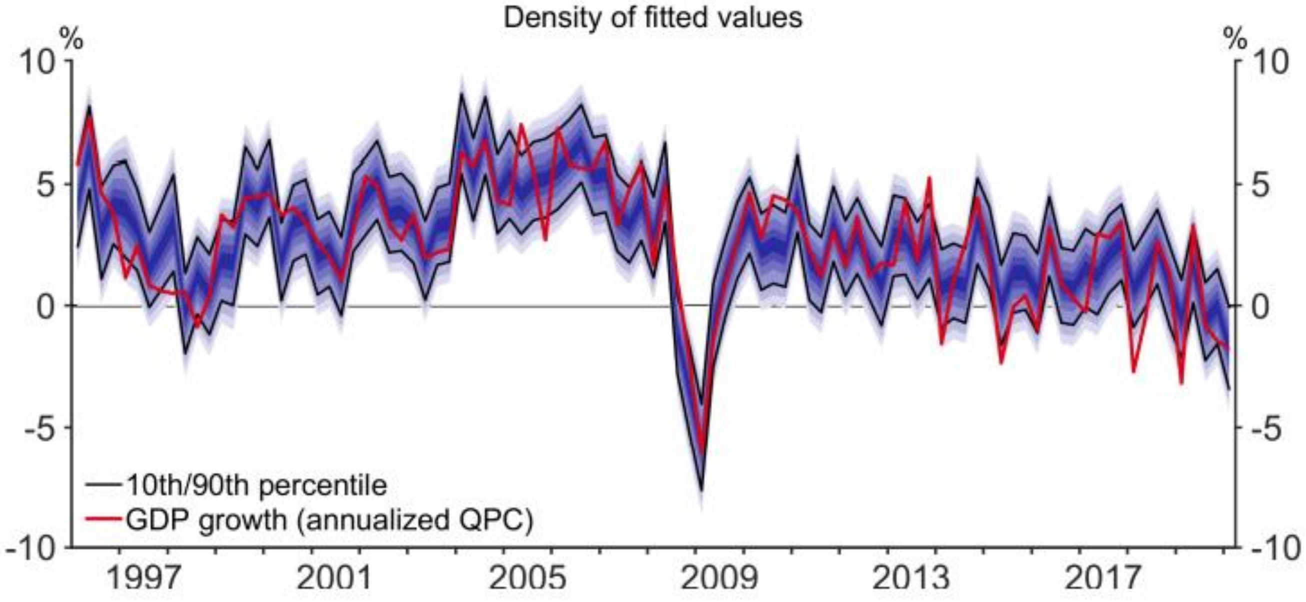
South African Reserve Bank Working Paper Series WP/21/01

Nowcasting South African GDP using a suite of statistical models

Byron Botha, Geordie Reid, Tim Olds, Daan Steenkamp and Rossouw van Jaarsveld

Authorised for distribution by Witness Simbanegavi

1 February 2021



Source: Botha, B., Olds, T., Reid, G., Steenkamp, D., & van Jaarsveld, R. (2021). Nowcasting South African gross domestic product using a suite of statistical models. South African Journal of Economics, 89(4), 526-554.

Variables Used:

1. Gross domestic product at market prices,
2. Long-term government bond yield,
3. Trade Activity Index, ABSA PMI, JP Morgan global manufacturing PMI,
4. Real effective exchange rate of the rand,
5. CPI, Value of building plans passed, Job advertisement space, Leading business cycle indicator, Industrial production, New Passenger Vehicle Sales, Electricity consumed in South Africa, Real Retail Trade Sales, M3, Physical volume of manufacturing production, and many more.


Discussion Paper 2019-03

RESERVE BANK OF NEW ZEALAND TE PŪTEA MATUA

Nowcasting GDP using machine learning algorithms: A real-time assessment

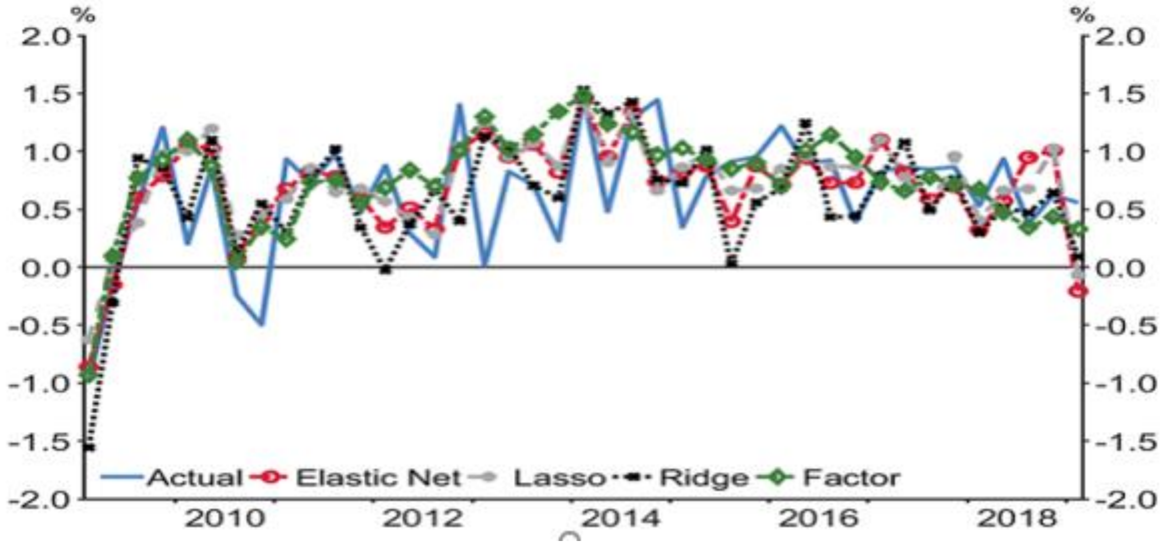
November 2019

Adam Richardson, Thomas van Florenstein Mulder & Tuğrul Vehbi

See The Journey of Te Pūtea Matua: Our Tāne Mahuta, available at: https://www.rbnz.govt.nz/about-us/the-journey-of-te-putea-matua-our-tane-mahuta

1. Several ML models were estimated over the 2009-2019 period using multiple vintages of historical GDP data and a large feature set comprising approximately 600 domestic and international variables.

2. The forecasts obtained from these ML algorithms were compared with the forecasting accuracy of a naive autoregressive benchmark, a dynamic factor model, and the official forecasts produced by the Reserve Bank of New Zealand.



Source: Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. International Journal of Forecasting, 37(2), 941-948.

Discussion Paper 2019-03

RESERVE BANK OF NEW ZEALAND
TE PŪTEA MATUA

Nowcasting GDP using machine learning algorithms: A real-time assessment

November 2019

Adam Richardson, Thomas van Florenstein Mulder & Tuğrul Vehbi

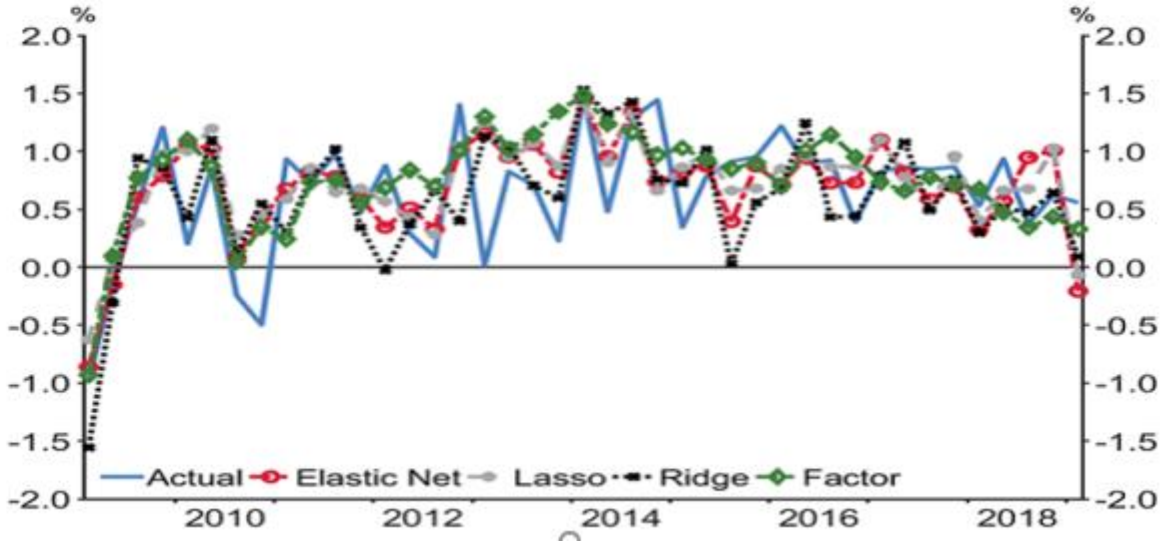
See *The Journey of Te Pūtea Matua: Our Tāne Mahuta*, available at:
https://www.rbnz.govt.nz/about-us/the-journey-of-te-putea-matua-our-tane-mahuta

Evaluated methods

1. Autoregressive Model (AR): This is a univariate AR model of order 1 for quarterly GDP growth.
2. Gradient Boosting: Specifically, the document mentions using the LS_TreeBoost algorithm. Gradient boosting is a method used to build a predictor from a series of individual weak models, often termed as 'learners'. The process involves fitting an initial model to the target variable and then adding subsequent models based on the residuals of the models fitted so far.
3. Support Vector Machine Regression (SVM): Mentioned in the context of its performance in nowcasting.
4. Neural Network: Also mentioned in the context of its performance in nowcasting.
5. Lasso Regression: A type of regression analysis that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model.
6. Elastic Net: A linear regression model trained with both L1 and L2-norm regularization.
7. Ridge Regression: A type of linear regression that includes L2 regularization.
8. Dynamic Factor Model: Used as a comparison to the machine learning models for nowcasting GDP.

The results indicated that the majority of the ML models produced point nowcasts that were superior to the simple AR benchmark. The top-performing models, which included boosted trees, support vector machine regression, and neural networks, could reduce average nowcast errors by approximately 20-23% relative to the AR benchmark.

Most of the ML algorithms also outperformed the dynamic factor model. Additionally, an average of the ML models appeared to add value to the official forecasts produced by the Reserve Bank of New Zealand.



Source: Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. International Journal of Forecasting, 37(2), 941-948.

Thank You