Report Title:

"Nowcasting GDP: A Comprehensive Analysis Using Alternative Data Sources"

Created by: Ali ElGamal

Date:29/10/2024

Documentation and Reporting for GDP Nowcasting Project

Introduction

With the ever-changing economic world, there is huge urgency for timely and accurate forecasting of GDP. Usual conventional methods rely on delayed data and cannot always capture the real-time variations in the economy. This use case project, therefore, investigates alternative data sources that could be integrated to support nowcasting of GDP, especially for Egypt. This effort leverages innovative data from Google Trends and mobility indicators, air quality indices, and high-frequency economic indicators to fashion a sound methodology for gauging the relevance and predictive power of these data sources.

This will focus on the alternative data sources that could be used, review methodologies for ranking, present a structured plan for data acquisition and their cost, relevance of each source to the specific economic conditions of Egypt. Hence, we look forward to making our contribution with such a systematic approach to deepening an understanding of economic dynamics in general and facilitating more informed decisions by policymakers and researchers.

Part A: Potential Alternative Data Sources

1. Data from Google Trends:

Google Trends provides the best insight into real-time economic activity through its monitoring of specific search terms. Its predictive power over various economic indicators, like unemployment rates, has been time and again proved. Multiple case studies underpin its potential use in nowcasting GDP; as such, it is one of the most important sources of data for this project.

2. Indicators of Mobility:

Data from platforms like Google Mobility Indicators have provided critical tools in measuring economic activities during the COVID-19 pandemic. These indicators offer a remarkably transparent picture of economic engagement and consumer behavior through tracking the movement into key locations like workplaces, retail spaces, and recreational sites.

3. Air Quality Indices:

The most recent analyses show that air quality data is a proxy for economic condition. Air quality often varies with the intensity of industrial activity; thus, it serves as an indirect means to perceive general economic well-being.

4. Surveys and Sentiment Indicators:

Business surveys and consumer sentiment indicators, which are conducted routinely, provide indications of economic expectations and confidence levels. This is important in understanding performance in the future about GDP.

5. High-Frequency Indicators of Economy

These leading indicators include a series of metrics such as industrial production, consumer surveys, and other timely data that can be aggregated to provide insight into GDP growth. The high-frequency indicators have been found quite useful in various economic contexts, especially in periods of volatility.

Case Studies from Other Countries

New Zealand: Applying machine-learning techniques to a rich dataset, both from domestic and international variables, is how scholars have tried to enhance forecasts on GDP growth. This methodology gave better results compared to the benchmark models, indicating how crucial the diversification of sources of information has been.

• Turkey: This study used, in Turkey, economic indicators together with machine learning methods. This approach tended to give lower forecast errors when the proposed technique was compared to a conventional one. Ensemble models, aggregating the predictions over several algorithms, further enhanced accuracy.

• Lebanon: Another example is a study using Google Trends and high-frequency indicators to nowcast, effective capturing of the GDP growth of Lebanon, hence proving flexibility in this data source for difficult economic situations.

• When focusing on Latin America, combining mobility indicators with Google search data within a standard modeling framework-like Dynamic Factor Models-added considerable value in studying the economic consequences of the COVID-19 pandemic.

Methodology for Ranking Alternative Data Sources

1. Identification of Alternative Data Sources
Begin by compiling a comprehensive list of potential alternative data sources identified through thorough research. Notable examples include:

- **Google Trends Data**
- **Air Quality Data**
- **High-Frequency Economic Indicators**
- **Social Media Sentiment Analysis**
- **Market Surveys**
- **Mobile Phone Usage Data**

2. Definition of Ranking Criteria
Establish clear criteria for evaluating and ranking each data source. Suggested criteria include:

- **Relevance: How highly it correlates the data source to GDP growth.**
- **Timeliness - How often is the indicator updated? How quickly does it reflect current economic conditions?**
- **Availability:** Accessibility of the data source within the target country, such as Egypt.
- **Quality:** Reliability and precision of the data - keeping in mind the possibility of bias.
- **Predictive Power:** Historical performance in forecasting GDP changes, potentially assessed through backtesting results.
- **Cost:** The financial implications associated with acquiring and processing the data.

3. Weighting the Criteria
Assign weights to each criterion according to its significance in the context of GDP nowcasting. For instance:

- **Relevance:** 30%
- **Timeliness:** 25%
- **Availability:** 20%
- **Quality:** 15%
- **Predictive Power:** 10%

4. Scoring Each Data Source
Evaluate each alternative data source against the established criteria using a scoring scale (e.g., 1 to 5, with 1 being poor and 5 being excellent). This evaluation should draw on literature reviews, expert insights, and empirical evidence.

5. Calculating Weighted Scores
Calculate the weighted score for each data source by multiplying the score for each criterion by its corresponding weight, then summing these values:
Total Score = $\sum(\text{Score}_i \times \text{Weight}_i)$

6. Ranking the Data Sources
Rank the alternative data sources from highest to lowest based on their total scores. The source with the highest score will be deemed the most effective for GDP nowcasting.

7. Model Validations and Sensitivity Analyses

Validation of rankings by comparing them with historical case studies or empirical results from other countries is necessary. Sensitivity analyses can be carried out in cases where any changes in weights or scores may affect the ranking to ensure the methodological robustness of the rankings.

8. Documentation and Reporting
Thoroughly document the methodology, including the rationale for the selection of criteria, scoring, and weighting. Present the findings in a clear format that highlights the top-ranked data sources and their potential applications for GDP nowcasting in Egypt.

Proposed Plan for the Selection of Top Three Data Sources

1. Google Trends Data

Data Accessibility and Collection Methods:

Google Trends data is available for anyone, either through the Google Trends website or API. One can search for any economic-related keyword, such as "unemployment" or "consumer spending", and get the time series data usually at a weekly frequency.

(ii) Acquisition Cost:

The good thing with Google Trends data is that access to it is free; hence, it is really an affordable resource for both researchers and policy operatives.

(iii) Relevance for Egypt:
This data offers insights into consumer behavior and sentiment in Egypt, particularly during economic fluctuations. Identifying keywords that reflect local economic conditions can enhance the data's relevance.
(iv) Limitations and possible biases:

• **Limitations:** Data may reflect only the search behavior of the internet users, which may not implicate the whole population. The seasonal events or media exposure may also affect trends.

• **Possible Biases:** Differences in Internet access between urban and rural areas or in the use of languages may make the sample not quite representative.

2. Air quality data

(i) Data Access and Collection Approach:

These data become available through governments, international organizations like WHO, and internet databases like OpenAQ. In general, the data are measured by monitoring stations that collect information on different types of pollutants, such as PM2.5 and NO2, and often provide real-time access via APIs.

(ii) Acquisition Cost:

Fortunately, most air quality datasets are publicly available at no charge. Specialized datasets or real-time monitoring services may be the exception, where several associated costs may arise.

(iii) Relevance for Egypt:

The amount of pollution is strongly associated with public health and economic productivity in Egypt; therefore, the report from this dataset is a key forward-moving indicator for economic activities across major urban cities.

(iv) Limitations and Potential Biases:

Limitation of Data: Typically, data is available with irregularities in geography-more monitoring in urban areas compared to rural areas. Historical data at times may be very scanty, which limits analysis over the longer term.

• **Potential biases:** Differences in the monitoring equipment and methods may create the measurement bias, while weather conditions changing the air quality will give rise to temporal bias.

3. High Frequency Economic Indicators

(i) Methodology to Access and Collect Data:

High-frequency indicators refer to different variables resulting from national statistical offices, central banks, and private organizations. Such examples of these include the industrial production indices, retail sales data, and consumer confidence surveys.

(ii) Acquisition Cost:

While most of the indicators can be freely available due to government sources, some proprietary sets may require subscriptions or even payments; the cost varies hugely depending on who the provider is.

(iii) Significance to Egypt:

These are the key indicators to expose short-run economic dynamics in Egypt, particularly in sensitive sectors such as manufacturing and retail.

(iv) Statement of limitations and prone biases:

• **Constraints:** Data may be revised; therefore, the reliability concerning forecasting could be compromised. Some of the indicators are of less desirable frequency and/or granularity.

• **Potential Biases:** The selection bias may come about if not all sectors are represented to the same degree. There is also reporting bias in cases where businesses are underreporting or overreporting, depending on how well they are doing economically.

Part B: Nowcasting GDP in Egypt Using Alternative Data Sources

2. Methodology

2.1 Data Collection

- **GDP Data**: Sourced from the Central Bank of Egypt and national statistical offices, covering the period from 2007 to 2021. This data includes quarterly GDP values.
- **Google Trends Data**: Selected keywords relevant to economic activities were tracked for search interest over the same period. Keywords included:
  - **Gold Prices**: "أسعار الذهب اليوم"
  - **Car Market**: "سوق السيارات

2.2 Data Preparation

Data Cleaning:

- **Identified and addressed missing values through interpolation methods.**
- **Data normalization was performed using Min-Max scaling to standardize the search interest scores to a common scale.**

2.3 Statistical Tools and Techniques

- Utilized **pandas** and **NumPy** for data manipulation.
- Employed **statsmodels** for regression analysis and statistical testing.
- Conducted Variance Inflation Factor (VIF) analysis to identify multicollinearity among features.

---

3. Model Development

3.1 Initial Models

- **Linear Regression**: Served as a baseline model to evaluate other approaches.
- **Support Vector Machine (SVM)**: Selected for its capability to model non-linear relationships through kernel functions.
- **Random Forest**: An ensemble learning method to improve prediction accuracy by combining multiple decision trees.
- **Elastic Net**: A regularization technique combining Lasso and Ridge regressions, effective for high-dimensional datasets.

3.2 Hyperparameter Tuning

Used **GridSearchCV** to fine-tune the SVM model, which involved a comprehensive search across multiple hyperparameters, including C, gamma, and kernel.

- **Best Parameters Identified**:
  - C: 100
  - gamma: 'scale'
  - kernel: 'linear'

3.3 Model Evaluation & back-testing

- **Metrics**:
  - **Mean Squared Error (MSE)**: Measures the average squared difference between observed and predicted values, reflecting model accuracy.
  - **R^2 Score**: Indicates the proportion of variance explained by the model, assessing overall model performance.

---

4. Results

4.1 Model Performance Summary

Model	Mean Squared Error	R^2 Score
Support Vector Machine (SVM)	222.47	0.89
Random Forest	289.32	0.86
Elastic Net	225.10	0.89

- **Best Performing Model**: Elastic Net achieved the lowest MSE and highest R^2 Score, demonstrating its effectiveness in capturing the relationships within the data.

---

5. Discussion

5.1 Interpretation of Results

The integration of Google Trends data into GDP prediction models showcased its potential as a leading indicator. The Elastic Net model, in particular, highlighted how search interest can correlate with economic activity, providing timely insights that traditional metrics may miss.

5.2 Visualizations

- Actual vs. Predicted GDP plots were generated to illustrate model performance visually, confirming the close alignment between predicted and actual values.

---

6. Conclusions

- The project successfully demonstrated that alternative data sources like Google Trends can enhance GDP nowcasting models.
- The consistent performance across multiple modeling techniques reflects the robustness of the findings.

7.References

1. https://www.imf.org/en/Publications/WP/Issues/2022/03/11/Nowcasting-GDP-A-Scalable-Approach-Using-DFM-Machine-Learning-and-Novel-Data-Applied-to-513703

2. https://dl.acm.org/doi/abs/10.1145/3524458.354725

3. https://www.boj.or.jp/en/research/wps_rev/wps_2022/wp22e09.htm

4.https://www.ibm.com/topics/multicollinearity#:~:text=The%20higher%20the%20VIF%20value,to%20ten%20indicates%20severe%20multicollinearity.

5. Anesti, N., S. Hayes, A. Moreira, and J. Tasker (2017). "Peering into the present: the Bank's approach to GDP nowcasting," Bank of England Quarterly Bulletin, Q2

6. Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018). "Macroeconomic nowcasting and forecasting with big data," Annual Review of Economics, 10, 615–643.

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