Nowcasting Macroeconomic Indicators using Google Trends

Master of Data Science Capstone
May 05, 2022

Submitted By:
Aishwarya Sharma
Harpreet Kaur
Jagdeep Brar

Project Partners:
Nick Newstead
Marina Smailes

Instructors:
Irene Vrbik
Firas Moosvi

Image Sources: UBC Logo, Statistics Canada Logo, Google Trends

Executive summary

The information on economic indicators is crucial for policymaking and taking decision at right time but this information is usually available with a lag. So, the need for alternative data sources is of growing importance for both supplementing Statistics Canada's data holdings and for nowcasting economic activity. The goal of this project is to develop a methodology to predict macroeconomic factors such as Gross Domestic Product (GDP), retail trade sales and retail e-commerce sales in real time by using the real time data source, Google Trends. Google Trends provides daily, weekly, and monthly reports on the volume of Google queries related to different industries which can help to understand the business cycles and provide signals about multiple aspects of the economy that can further be used to estimate the economic factors in real time. The nowcasting of economic indicators will provide more timely information for policymaking.

1. Introduction

Macroeconomic factors are the key drivers of economy, and their timely information helps in good policymaking. However, this information is available with a lag, for instance, the data for the present month's GDP is generally published in the coming month/quarter which causes delay in decision-making. To overcome this issue of delayed information gave rise to nowcasting approach. This approach hasrecently gained the interest of economists and researchers as this approach provides the information on economic indicators in real-time. Traditional macroeconomic indicators have some lag, and to fill this gap of information, Google Trends have been widely used as it may help in predicting the present [1]. The volume of queries on different industries may be correlated with the current level of economic activities in respective industry and may help to predict the subsequent data release [1].

Many researchers have used Google trends for nowcasting the economic activity. Google Trends provide information of business cycles and economic activities in economy and the salient features of these business cycles can be captured with few unknown factors using dynamic factor analysis models [2]. These models are applicable to high-dimensional data and can reduce the dimensionality of economic systems. Dynamic Factor Model (DFM) became the mainstream tool for nowcasting GDP growth over the time. Later on, new techniques emerged, and researchers have started to use machine learning algorithms for nowcasting economic factors. Woloszko [3] proposed a weekly tracker to estimate GDP in 46 Organisation for Economic Co-operation and Development (OECD) countries and G20 countries (excluding European Union). The proposed OECD tracker is based on a machine learning algorithm that estimates the relationship between Google Trends variables and GDP growth.

Dauphin et al. [4] have also used Google Trends data to estimate GDP growth, they provide comparative analysis of different nowcasting approaches such as Auto-Regressive (AR) models, DFM and some machine learning algorithms like Regularized Regression models, Random Forest, Support Vector Machine (SVM) and Neural Networks, and state that there is no one-size fits all model as different models are suitable for different datasets. Richardson et al. [5] used machine learning algorithms to nowcast GDP growth in New Zealand and their results show that machine learning algorithms boosted trees, SVM and neural networks outperformed the traditional AR models for their study. The aforementioned studies indicate that traditional econometrics models and machine learning models both can be used for the nowcasting economic factors,but the success and accuracy of the model may vary for different datasets. Therefore, a comparative study between traditional and modern machine learning algorithms may be more appropriate to fit a model on data in hand.

2. Aims and Objectives

The aim of the project is to develop a methodology to predict macroeconomic indicators such as GDP, retail trade sales and retail e-commerce sales with real-time data source, **Google Trends**. The volume of queries for different keywords and categories from Google Trends API will serve as the predictors for nowcasting the desired economic factors. The key goals of the project are discussed below:

1) **Nowcasting quarterly GDP**: Our first goal is to nowcast the macroeconomic indicator GDP quarterly at national level by using the real time Google Trends variables as predictors.
2) **Nowcasting monthly retail trade sales**: The retail sales data are available monthly, so our objective is to nowcast the monthly retail trade sales at national level**.**
3) **Nowcasting retail e-commerce sales**: The retail sales data are available monthly, so our objective is to nowcast the monthly retail e-commerce sales at national level.
4) **Nowcasting economic indicators at industry level**: If time permits, we aim to nowcast quarterly GDP, monthly retail trade sales and monthly e-commerce retail sales at the industry level for a selection of key industries. The data for GDP is available quarterly and monthly, so we will try to nowcast monthly GDP at national level.

3. Data Set

Data set for this project are open ended and the short description about data is provided below:

1) **Gross Domestic Product (GDP) at basic prices monthly:** This dataset is a comma separated file containing the information about the monthly GDP.

This file contains data from 1997 and do have some missing values, thus will require data wrangling.

2) **Gross Domestic Product (GDP) at basic prices quarterly**: This dataset is comma separated file containing the information about the GDP quarterly. This file contains data from 1997 and also requires data wrangling.
3) **Gross domestic product (GDP), expenditure-based, Canada, quarterly**: This dataset is comma separated file containing the information about the GDP quarterly. This file contains data from 1997 and do have some missing values, thus will require data wrangling.
4) **Retail trade sales by province and territory**: This dataset contains information about the retail sales as per the province and territory. This data file is also comma separated and will require data wrangling.
5) **Retail trade sales by industry**: This is a comma separated data set containing the information about the retail sales trades as per the industry. Data wrangling is required in this dataset as well.
6) **Retail sales, price, and volume**: This is a comma separated data set containing monthly retail sales, price, and volume data. This data set will need some data wrangling.
7) **Retail E-commerce sales**: This is a comma separated dataset containing the information about the retail e-commerce sales. This is the data for digital sales and will require some data wrangling as well.
8) **Google Trends API**: we will be accessing Google Trends website to get real time data for the macroeconomic indicators.Different keywords, categories and subcategories will be used to extract Google Trends predictors such as Economic crisis, loans, GPS, unemployment, affordable housing, economy news, agriculture, and forestry [3, Annexure B].

Our focus will be on the data starting from 2004 for GDP and retail trade sales data as we have Google trends available from that period. The e-commerce data is available from the year 2016. This range of data will provide us enough information for nowcasting.

4. Methodology

The detailed description on the methodology of the whole project is as mentioned below and the workflow is also depicted in Figure 1.

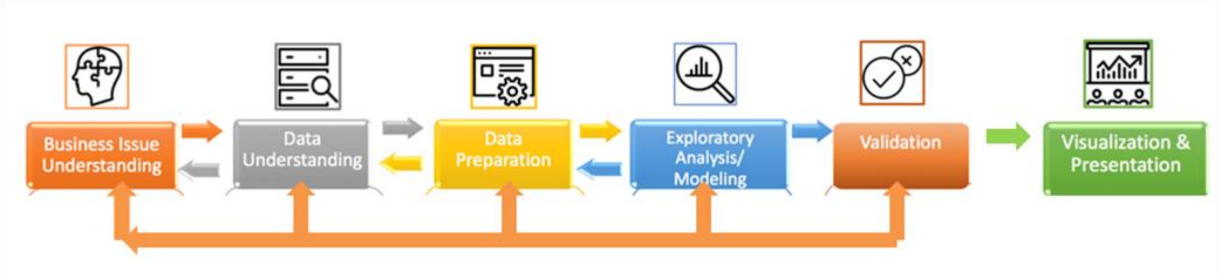

Figure 1. Workflow of the project

(i) **Data Cleaning and Wrangling**

The dataset that we have mentioned above for the GDP contains the information about Canada and different industries. The dataset for GDP, retail sales and e-commerce sales will be filtered so that it contains information starting from the desired year. Moreover, the data are time series and need check for cycling patterns, seasonality and volatility. The time series may need transformations in order to get stationary time series that can be further used for model fitting. In addition to that, Google Trends data are also time series data and need to be checked for seasonality patterns. There will be need to change the frequency of Google Trends data so that it matches with the Statistics Canada's data frequency. Therefore, time series may require many transformations.

(ii) **Model Fitting**

The data set is a time series data and will require the nowcasting using the real time data from Google Trends. The volume of real time keyword queries from the Google Trends will be used as the predictors to nowcast the macroeconomic indicators. As mentioned in Section 1, the econometric models and machine learning models both can be used to estimate the relationship between predictors and macroeconomic factors. Following the similar path, we intend to use methods of both categories to fit our data. The intended methods are discussed below:

- **Econometric models**: Since the data is time series and DFM have been widely used for such analysis, we intend to use DFM which will serve as a dimension reduction method and can be helpful in analysing the time series data. Also, we are aiming at using AR models to nowcast the indicators. These models may provide us good fit.
- **Machine learning models**: The literature claims that machine learning models have been serving well to estimate abrupt changes in economic activity. So, we intend to use LASSO, Random Forest and Neural Networks depending upon the time constraints.

(iii) **Validation**: We will be implementing the above-mentioned models and will be performing the comparative analysis. This will provide us the accurate predictions and will let us choose the best model for nowcasting economic indicators.

5. Deliverables and timeline

For our project we would be having following deliverables:

(i) **Visualizations**

- Once the model fitting is achieved, predictions of GDP (quarterly), monthly retail revenue and monthly digital sales will be presented in a tabular format.

- This will be followed by bar graphs and line charts for the demonstration of the predictions.
- A dashboard will be created which will give the information about the nowcasting of the indicators.

(ii) **Final Presentation and Report**

- A final presentation deck will be created which will highlight the research results and predictions.
- This will be structured and concise.
- Also, a report will be created that documents the methodology and findings of the research results.

These outlined deliverables will be achieved with the pre-assumed timeline, as mentioned below:

Table 1. Timeline of the project

Phases and Weeks	1st week	2nd week	3rd week	4th week	5th week	6th week
Project and Data Understanding & Proposal						
Data Preparation and Exploratory Analysis						
Modelling and Validation						
Comparative/More granular Study						
Prepare Dashboard, Report, and Presentation						
Research						

We will start the project as soon as the first meeting kicks off. The total time taken for the final submission of the project would be in 7 weeks, the final presentation would be delivered on June 22, 2022.

Week 1: Project and Data Understanding & Proposal

- Discussion with project partners to define scope of work – 2 days
- Literature review – 1 days
- Proposal document and review – 2 days

Week 2:Data Preparation and Exploratory Analysis

- Clean the dataset and extract the useful factors–1 day
- Handle the missing values using interpolation of the data – 1 day
- Standardization and Transformation might be applied as required – 2 days
- Exploratory analysis – 1 day
- Follow-up meeting for review

Week 3 & 4: Modelling and Validation

- Exploring Time series with machine learning models–4 days
- Fit multiple models – 4 days
- Testing and validating the models and selecting the best one – 2 days
- Follow-up meetings for review

Week 5: **Comparative/More granular Study**

- Comparative study of econometrics and machine learning models – 2 days
- Resolve if there are any issues in model fitting – 2 days
- Now monthly GDP if time permits – 1 day

Week 6: **Prepare Dashboard, Report, and Presentation**

- Prepare dashboard – 2 days
- Report writing – 2 days
- Presentation – 2 days

We will be aiming at using Python for all our coding, model fitting and analysis. Also, Git Hub will be the channel where we will be committing our code.

6. Roles and Responsibilities

We as a team have decided to accomplish our project with equal distribution of responsibilities. Our internal team coordination and communication would be crystal clear. The phase of data wrangling for the Google trend data and the Statistics Canada data would be distributed amongst all the members as per the size and work required in every file. The phase of modelling would be distributed amongst the team individuals. Every member will independently perform the allocated modelling techniques. However, later on the results would be combined and presented.

7. Conclusion

This project will provide us the timely information on macroeconomic indicators that would be helpful for policy making. The real time data of Google Trends will be used to nowcast GDP, retail trade sales, retail e-commerce sales. The visualisation tool will help to observe the macroeconomic factors of different industries, and this would be user friendly and interactive.

References

[1] H. Choi, H. Varian, Predicting the present with Google Trends, *Economic record*, *88 (2012)*, 2-9.

[2] Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415-525). Elsevier.

[3] Woloszko, N. (2020). Tracking activity in real time with Google Trends, OECD Economics Department Working Papers, No. 1634, OECD Publishing, Paris.
[4] Dauphin, M.J.F., Dybczak, M.K., Maneely, M., Sanjani, M.T., Suphaphiphat, M.N., Wang, Y. and Zhang, H., 2022. *Nowcasting GDP-A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies*. International Monetary Fund.
[5] Richardson, A., van Florenstein Mulder, T. and Vehbi, T., 2021. Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), pp.941-948.