**TRACKING GROWTH DURING THE NEW WAVE OF COVID-19 – USING GOOGLE TRENDS**

**Abstract:**

The immediate impact of coronavirus requires the high-frequency data to adapt well to the variation of the macro-economy. Applying the Artificial Neural Network algorithm, this thesis aims to construct a Weekly Tracker, which is used for tracking the growth rate, from the Google Trends dataset. Compared to the Autoregressive model, this thesis shows that predictions from the Artificial Neural Network algorithm outperform in leading information of growth rate movement. The Weekly Tracker suggests that the recovery of 22 economies in the sample has been constrained during the new wave of Covid-19. The growth rate estimate may be still positive, but the trend is horizontal, even in developed countries. And with the Shapley value analysis, this thesis finds out Consumption topics and categories, especially trade and services, are the variables that have a key role for the recovery; Economic Anxiety topics and categories are extremely popular during the coronavirus spread out, especially in the beginning, and have large negative contributions; number searches of Labor Market and Health Condition topics and categories also increase during the pandemic and they contribute positively to the Weekly Tracker. Focusing on Vietnam, this thesis finds out the down trend of the Weekly Tracker during the social-distancing and locked down period in Ho Chi Minh City, same as the movement of Consumption searches number.

Contents

[CHAPTER 1: INTRODUCTION 1](#_Toc90665587)

[1.1 Problem statement 1](#_Toc90665588)

[1.2 Research approach 2](#_Toc90665589)

[1.3 Research objectives 3](#_Toc90665590)

[1.4 Research questions 4](#_Toc90665591)

[1.5 Thesis outline 4](#_Toc90665592)

[CHAPTER 2: LITERATURE REVIEW 6](#_Toc90665593)

[2.1 The potential of the Google Trends data 6](#_Toc90665594)

[2.1.1 What is Google Trends? 6](#_Toc90665595)

[2.1.2 The application of Google Trends data 9](#_Toc90665596)

[2.2 The application of Google Trends in Covid-19 11](#_Toc90665597)

[2.3 Some concerns about Google Trends data 15](#_Toc90665598)

[2.4 Summary 16](#_Toc90665599)

[CHAPTER 3: DATA AND METHODOLOGY 17](#_Toc90665600)

[3.1 Model 17](#_Toc90665601)

[3.2 Estimation Methods 18](#_Toc90665602)

[3.2.1 Artificial Neural Networks algorithm 18](#_Toc90665603)

[3.2.2 SHAP tool 20](#_Toc90665604)

[3.3 Data 22](#_Toc90665605)

[3.3.1 Outcome variable 22](#_Toc90665606)

[3.3.2 Google Trends variables 22](#_Toc90665607)

[CHAPTER 4: RESULTS 24](#_Toc90665608)

[4.1 Outcome from ANN algorithm 24](#_Toc90665609)

[4.1.1 Performance of the output from the training step 24](#_Toc90665610)

[4.1.2 The contribution of variables to the outcome 28](#_Toc90665611)

[4.2 Weekly analysis 30](#_Toc90665612)

[4.2.1 The fluctuation of Weekly Tracker during the new wave of Covid-19 30](#_Toc90665613)

[4.2.2 The positive signal at the end of 2021 36](#_Toc90665614)

[4.2.3 The factors having large contribution to the growth rate during the Covid-19 pandemic 37](#_Toc90665615)

[4.3 Discussion 39](#_Toc90665616)

[4.3.1 The expectation of the recovery at the end of 2021 39](#_Toc90665617)

[4.3.2 The case of Vietnam 42](#_Toc90665618)

[CHAPTER 5: CONCLUSIONS AND POLICY IMPLICATIONS 45](#_Toc90665619)

[5.1 Conclusions 45](#_Toc90665620)

[5.2 Policy implications 46](#_Toc90665621)

[5.3 Limitations 48](#_Toc90665622)

[5.4 Recommendations for future work on this topic 48](#_Toc90665623)

[REFERENCES: 49](#_Toc90665624)

[APPENDIX 52](#_Toc90665625)

**LIST OF TABLES**

[Table 2.1 The application of Google Trends data 10](#_Toc90637751)

[Table 2.2 The need for big data in the Covid-19 pandemic research 12](#_Toc90637752)

[Table 4.1 Forecast performance, comparing to AR(4) model 26](#_Toc90637753)

[Table 4.2 Top categories and topics since the coronavirus spreads out 41](#_Toc90637754)

**LIST OF FIGURES**

[Figure 1.1 Quarterly GDP growth of 22 Asia and The Pacific 2](#_Toc90637929)

[Figure 3.1 The model of a neuron 19](file:///C:\Users\MyPC\Desktop\VoPhuocTri_FourthDraft%20-.docx#_Toc90637930)

[Figure 3.2: Multilayer feedforward architecture 20](file:///C:\Users\MyPC\Desktop\VoPhuocTri_FourthDraft%20-.docx#_Toc90637931)

[Figure 4.1 Outcome from model and actually quarter GDP growth rate. 25](#_Toc90637932)

[Figure 4.2 Tracker's prediction for 2021 Q2 and Q3 27](#_Toc90637933)

[Figure 4.3 Most important variables and their contributions to predictions 28](#_Toc90637934)

[Figure 4.4 A focus on 2021 Q2 29](#_Toc90637935)

[Figure 4.5 Weekly Tracker: High income economies 32](#_Toc90637936)

[Figure 4.6 Weekly Tracker: Upper-middle income economies 34](#_Toc90637937)

[Figure 4.7 Weekly Tracker: Upper-middle income economies 36](#_Toc90637938)

[Figure 4.8 Most recent predictions of the Weekly Tracker 37](#_Toc90637939)

[Figure 4.9 Contribution of variables to predictions during the Covid-19 pandemic 38](#_Toc90637940)

[Figure 4.10 Drivers of the growth rate during the Covid-19 new wave 39](#_Toc90637941)

[Figure 4.11The Weekly Tracker of Vietnam during the new wave of Covid-19 43](#_Toc90637942)

[Figure 4.12 Drivers of the growth rate during the Covid-19 new wave in Vietnam 44](#_Toc90637943)

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **GDP** | **Gross Domestic Product** |
| **SVI** | **Search volume index** |
| **ANN** | **Artificial neural network** |
| **AR** | **Autoregressive** |
| **RMSE** | **Root Mean Squared Error** |
| **Covid-19** | **Coronavirus Disease Of 2019** |
| **RMSFE** | **Root mean squared forecast errors** |
| **PMI** | **Purchasing Managers' Index** |
| **FA-MIDAS** | **Factor-Augmented Mixed Data Sampling** |
| **Q** | **Quarter** |
| **OECD** | **Organisation for Economic Co-operation and Development** |
| **ADB** | **Asian Development Bank** |
| **GSO** | **General Statistics Office of Vietnam** |
| **WHO** | **World Health Organization** |

**LIST OF COUNTRIES ABBREVIATIONS**

|  |  |
| --- | --- |
| **AM** | **Armenia** |
| **AU** | **Australia** |
| **AZ** | **Azerbaijan** |
| **BN** | **Brunei Darussalam** |
| **CN** | **China, People's Rep of** |
| **GE** | **Georgia** |
| **HK** | **Hong Kong, China** |
| **IN** | **India** |
| **ID** | **Indonesia** |
| **JP** | **Japan** |
| **KZ** | **Kazakhstan** |
| **KR** | **Korea, Republic of** |
| **KG** | **Kyrgyz Republic** |
| **MY** | **Malaysia** |
| **MN** | **Mongolia** |
| **NZ** | **New Zealand** |
| **PH** | **Philippines** |
| **SG** | **Singapore** |
| **LK** | **Sri Lanka** |
| **TW** | **Taipei,China** |
| **TH** | **Thailand** |
| **VN** | **Vietnam** |

# INTRODUCTION

## Problem statement

In 2021, the economic growth of Asia and Pacific countries, after recovering in the first haft, have slightly losses in the third quarter during the new wave of the Covid-19 pandemic. According to the ADB September update report, the growth rate of 2021 is forecasted down from 7.3% to 7.1% for Asia developing, 4.4% to 3.0% for Southeast Asia, 9.5% to 8.8% for South Asia, 1.4 to -0.6% for The Pacific. Only Central Asia and East Asia are noted that the maintenances of the recovery (up to 4.1 for Central Asia and 7.6% for East Asia)[[1]](#footnote-1). Figure 1.1 points out the expectation flop of growth rate in the third quarter of 2021 in Asia and The Pacific countries, although they get the recovery during the first haft of 2021. China, Vietnam, Korea, and Taiwan – countries that already have reported GDP growth rate of the third quarter – acknowledge the downward since the new wave of the Covid-19 pandemic. The wave of Covid-19 pandemic breaking in April, with a new fast-spreading variant, has seriously impacted the health and lives of people and interrupted the production and business operations (GSO, 2021). The immediate impact of coronavirus requires the high-frequency data to adapt well to the variation of economic activity.

For avoiding the spread of the virus, multiple protocols such as reducing travel, social distancing, and even lockdown are promulgated. It is a barrier for the researcher to conduct surveys, which is a popular method to collect data. Moreover, the dataset from official sources is limitative in the situation of the pandemic, it is concerned about the lag problem, which reduces the relevance to the real-time situation and leads to mistakes in predictions. Meanwhile, with the variety of categories and topics, the length of time, and the large scale of countries, the Google Trends dataset probably reacts well to the rapid rising and falling of the economic activity in the epidemic period (Monache et al., 2021, Jardet & Meunier, 2020; Lewis et al., 2020; Woloszko, 2020; Tamara et al., 2020; Zhemkov, 2021; Ankargren & Lindholm, 2021; Eraslan & Gotz, 2020). In the addition, the diversity of categories and topics of Google Trends provides the tendencies to expand the width of the dataset, which is considered the approach to come closer to the best prediction result (Lewis et al., 2020). Not only the issues of the protocols of contact restricting but also the expensive cost and scoped limitation, traditional surveys are disadvantage during the pandemic when combining datasets from Google. With just a computer, the researcher has the availability to get data of any country at any time (as long as after 2004 – the year that Google Trends calculates the index) with multiple frequencies. The potential of the Google Trends data is the motivation of this thesis.

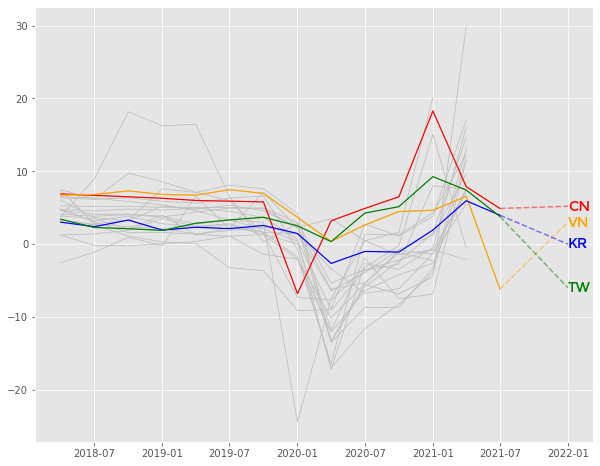


Figure . Quarterly GDP growth of 22 Asia and The Pacific Countries  
Note: Highlighted countries have officially reported GDP growth rate in the third quarter of 2021, while others only have the second quarter.  
Source: ADB data

## Research approach

This thesis aims to construct a Weekly Tracker that tracks the GDP growth rate based on Google Trends data and countries' dummy variables. The index will display the situation of 22 economies from Asia and The Pacific. However, the application of Google Trends data faces a challenge about the estimation methods to deal with the large dataset. Giannone et al. (2018) highlighted the requirement of carefulness when using big data to forecast or nowcast, the parametric estimations are concerned with providing erroneous predictions. It means that the research applying big data is in the need of new methods and machine learning algorithms are the most potential options. In recent years, machine learning algorithms are largely put in the economic research of many organizations (World Bank, OECD, ADB, etc.) or national banks (Bank of England, Bank of Germany, etc.).

Following Woloszko (2020), this thesis applies the two-step model. In the first step, for trainning the computer how to apply the algorithm and the machine learning model on the dataset to predict the outcome, training and testing proceeds for the dataset, which includes quarterly Google Trends and quarterly GDP growth rate, in the period from the first quarter of 2005 to the second quarter of 2021. And in the second step, a Weekly Tracker will be constructed by nowcasting from weekly Google Trends data to study the economic situation during the third quarter and first haft of the fourth quarter of 2021.

Artificial Neural Network algorithm is applied to train the dataset in this thesis. In recent years, this method is increasingly popularly used in economic research (Jene et al., 2020; Sanusi et al., 2020; Malte, 2018; Loerman & Benedikt, 2019; Woloszko, 2020). The dataset of this thesis includes around 86 categories and 75 topics from Google, so it is needed to consider the overfitting problem. But by the multiple-layer structure of the ANN, this problem is able to be avoided. After using the ANN algorithm to train the dataset, the predictions are tested the valuable by comparing to the Autoregressive lag 4 model in Root Mean Squared Error. Eventually, this thesis takes "a dive into the black box" to investigate how the predictors relate to economic activity. The SHAP tool is leveraged to interpret the relationship of the independent variables to the predictions. This method combines the efficiency and symmetry properties, and it is considered to have the power to provide explaining given predictions (local interpretability) or the general functioning of a model (global interpretability) (Woloszko, 2021; Molnar, 2020). This insight will be analyzed in more detail in Chapter 3.

## Research objectives

This thesis aims to construct a reliable Weekly Tracker, by applying Google search data and machine learning, to track the GDP growth rate during the Covid-19 pandemic. After using the two-step model and Artificial Neural Network algorithm, the first objective of this thesis is to prove that the output of the model is valuable for tracking the GDP growth rate in the sample of the training dataset, by comparing it to the result of the AR(4) model in RSMEs.

After proving the trustworthiness of the outcome of the model, the second objective is to build up a Weekly Tracker and study how it fluctuates during the Covid-19 pandemic new wave in the third quarter and the first haft of the fourth quarter 2021, in 22 Asia and The Pacific countries.

Finally, with the objective of contributing to the policy, this thesis will apply the SHAP tool to find out which variables have large contributions to the prediction output, thereby we can know which factors play key roles for GDP growth rate during the fourth wave of Covid-19.

## Research questions

Based on the objectives presented above, this work has two main research questions:

1) Does the Tracker built up from this thesis perform reliably in order to track GDP growth rate of Asia and Pacific economies?

2) How Weekly Tracker fluctuates during the new wave of the Covid-19 pandemic in Asia and Pacific economies?

3) Which factors play key roles for GDP growth rate during the new wave of the Covid-19 pandemic in Asia and Pacific economies?

## Thesis outline

The structure of this thesis is shown as follows. This chapter is the Introduction. So far, it provides an overview of the topic, the approach, the objectives, and the main questions of this work.

Chapter 2 covers the theoretical framework and literature review. It will provide the basic information about Google Trends data: how to collect data, how to calculate the index, and the structure of the dataset. The vision of the potential of Google Trends data will be presented, as well the application of this dataset in previous empirical works. Chapter 2 will present multiple papers about applying Google Trends data for forecasting and nowcasting economic growth in the Covid-19 pandemic is reviewed. On the other hand, Chapter 2 will deliver some concerns the value of Google Trends in economic research.

Chapter 3 will furnish the methodology. Firstly, it will show the two-step model (how to build the model, and how to apply it). Secondly, the dataset detailed information will be provided: the source of data, the scope of the dataset, the preprocessing data, and the final standardization. And Chapter 3 also introduces the Artificial Neural Network algorithm and interpretability SHAP tool.

Chapter 4 will show the result. The valuation of the predictions will be tested by RMSEs. After that, the result of the Shapely values will inform how the weekly Google Trends series contribute to the outcome of model. The movement of Weekly Tracker will be presented during the third quarter and the fourth quarter first haft of 2021. And the case of Vietnam will be discussed in this Chapter.

Chapter 5 will provide the conclusions, policy implications and the contributions and limitations of this thesis.

# LITERATURE REVIEW

This chapter provides the Google Trends data and tracking economic activity in the Covid-19 pandemic literature. Because this thesis purposes to point out the direction to construct the index that is able to react timely with the fluctuation situation practically, this chapter tries to concentrate on empirical literature.

Nowadays, the impacts of the Covid-19 pandemic on economies are extremely rapid and significant. The spread of the coronavirus is one of the biggest challenges that the governments of all countries around the world have faced since early 2020. It leads to the requirement of timely reacted research to the real-time situation. However, even at the beginning of the epidemic, traditional data showed the limitation to help researchers meet this demand. Firstly, the official data, which is popular in research, is considered poor in reacting to real-time fluctuation (Woloszko, 2020). Secondly, because of the worry about the spreading of the virus, the government of many countries promulgates multiple protocols reducing travel, social distancing, and even lockdown. It is difficult for the researcher to conduct surveys, which is another popular source and also is higher than official data source in the level of reacting real-time situations.

Multiple working paper series in 2020 and early 2021 suggested the need for new higher frequency data, particularly Google Trends, in the economic field. The Google Trends data is used in research for a long time, in the large area of fields, and in recent years, there is progress in applying it in studying several economic factors. This chapter is constructed with two sections. The first section concentrates on the Google Trends data. It will introduce the basic information about the data source (how it is collected, calculated, and what the index represents), the manners of application the Google Trends data, and some empirical research applying Google Trends. The second section presents the valuation of Google Trends data for the economic predictions during the Covid-19 pandemic.

## The potential of the Google Trends data

### What is Google Trends?

Google Trends is a tool provided by Google. Google Trends reports an index of search activity. Google Trends supports users with two options, one is time-series and the other is geographical location. In both two options, there are three steps to measure the index. In the first step for time-series option, we choose a query that relates to the question we concern about and a period at a particular area. In the second step, Google measures the proportion of the number search of queries including the question in total queries in the chosen period. If the period is less than three months, the result is daily. If the period is larger than three months and smaller than five months, the result is weekly. And if the period is larger than five years, the result is monthly. In the third step, Google ranks levels of proportion by day or week, the index is calculated by dividing daily/weekly proportion by the highest day/week proportion and time to 100. It means, according to Google definition[[2]](#footnote-2): “A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.” In the geographical location option, three steps are similar. Firstly, we choose a query about the question we concern about and geography at a particular time. In the second step, Google measures the proportion of the number search of queries including the question in total queries in chosen geography. In the third step, Google ranks levels of proportion by location, the index is calculated by dividing location proportion by the highest location proportion and time to 100. And also, according to Google's definition, we have: “100 is the location with the most popularity as a fraction of total searches in that location, a value of 50 indicates a location which is half as popular. A value of 0 indicates a location where there was not enough data for this term.”

**The query can be specific keyword, topic, or categories. Generally, topic and categories have been more favored than specific keyword recently. The reason is specific keyword may lead to ambiguity problem (Woloszko, 2020). For example,** the "Apple" keyword is related to not only the fruit but also the company. Or other instance is point out by Baker & Fradkin (2013), they need "jobs" keyword to measure job search, however, they realized that the keyword includes the term of personal name, "Steve Jobs". Some particular keywords have multiple meanings, thereby it leads to the risk that the researcher probably chooses the wrong term, and it is dangerous for the research (Stephens-Davidowitz & Varian, 2014). **Meanwhile,** the topics are created by combining multiple requests made on Google stand on their purpose and meaning and taking into account where users click (Woloszko, 2020). It allows to single out searches related to the term directed by the researcher and removes the keywords that belong to other terms. For example, if we choose the topic "Apple (company)", it combines searches for keywords such as "iPhone", "iPad", "apple watch", and "MacBook"; and it understandably excludes fruit term.

Google classifies search queries into about 1200 categories[[3]](#footnote-3). They are constructed as a 5-level hierarchical classification and aggregate together all searches related to each of them. The categories are assigned to individual searches by a probabilistic algorithm (Varian & Choi, 2011). For example, the queries would be allocated to the category **Vehicle Tires** which is a subcategory of **Auto Parts** which is a subcategory of **Automotive**. Another reason helping topics and categories outperformance specific keywords is they are harmonized across language (Woloszko, 2020). For instance, in term of a specific keyword, if we want to investigate the keyword "Birthday", we have to input "Sinh nhật" for Vietnamese, or "お誕生日**" for Japanese; meanwhile, the topic "Birthday" is similar in term of both Vietnamese and Japanese, and other languages. The categories, which are included in the Google Trends dataset, perform the same as the topic. With the number of topics and categories, the Google Trends dataset is believed that tends to cover a large area of economic fields (Stephens-Davidowitz & Varian, 2014). Furthermore, the data has been collected since 2004, and from all countries using Google, it has the ability to provides a large panel dataset for research activity (Woloszko, 2020). Therefore, there is an advancement in the implementation of the Google Trends index in research.**

**However, there are some concerns warned by literature when applying the Google Trends index. The first is careful when interpreting the trends in search behavior in the case of long-term (**Stephens-Davidowitz & Varian, 2014)**. For instance, U.S. searches related to the keyword "science" are downward since 2004. It does not mean that the interest in science declines. The reason is the Google searchers' composition has changed through time. In 2004, colleges and universities where searches on science were popular used the internet heavily. At the present, the composition of internet users has become broader. Chapter 3 will present how to resolve this problem. The second is watchful in conclusions based on the relative value of two searches at the national level (**Stephens-Davidowitz & Varian, 2014). For instance, in the U.S., the number of searches related to "Jewish" is 3.2 times more than "Mormon". It is not able to conclude that the Jewish population is 3.2 times larger than the Mormon population. It may be the Jewish people uses the internet in higher proportions than the Mormon people. Generally, the Google Trends data is more valuable for relative comparisons.

**So far, this section has already introduced some basic information about Google Trends. It presented what Google Trends is, what the Google Trends index stands for, how the data is collected and calculated. Furthermore, it shows the reason why recent research prefers topics and categories to a specific keyword, thereby pointing out the advantage of the Google Trends data application. On the other hand, potential pitfalls are warned. Next, several empirical research applying Google Trends will be provided to show the potential of this dataset. These papers were written before 2020, my purpose is to point out the large area fields that Google Trends can cover. The application of this dataset for the Covid-19 pandemic will be highlighted in section 2.2.**

### The application of Google Trends data

The fields applied Google search dataset is broad. In recent years, Google Trends is increasingly applied in multiple economic fields of research, such as unemployment (Baker & Fradkin, 2017), private consumption (Vosen & Schmidt, 2011), recession (Chen et al., 2014), consumer confidence (Choi & Varian, 2011), finance (Balakrishnan & Kalpit Dixit, 2013), and even grain prices (Martinez et al., 2011). Table 2.1 provides some popular papers applying the Google search index in the economic field and shows their insights into the valuation of this dataset. It is widely believed that the Google Trends index is outstanding in covering economic sectors because of its large scope, and it is also advantageous in predicting activity.

Table . The application of Google Trends data

|  |  |  |  |
| --- | --- | --- | --- |
| **Paper** | **Sector** | **Keyword** | **Finding** |
| Choi & Varian (2011) | Motor vehicles and parts, Initial claims for the unemployment benefits, Travel, Consumer confidence | "Trucks & SUVs" and "Automotive Insurance"; "Local/Jobs" and "Society/Social Services/Welfare & Unemployment"; "Hong Kong" under the category "Vacation Destinations"; "Trucks & SUVs", and "Hybrid & Alternative Vehicles" | They do not believe that Google Trends is valuable for *predicting the future*, but it is really helpful for *predicting the present* |
| Suhoy (2009) | Industrial production, Retail trade, Revenue of trade and services, Consumer imports, Exports of services, and The employment rate | 31 queries follows catefories which are "Automotive", "Business", "Home & Garden", "Beauty & Personal Care", "Food & Drink", "Travel", "Real Estate", "Shopping", "Industries", "Computers & Electronics", and "Finance & Insurance" | Query categories are valuable for growth-cycle prediction, especially these subcategories: "Recruiting and Staffing", "home appliances", "Travel", "Real estate", "Food and Drink", and "Beauty and Personal care". |
| Chen et al. (2014) | Fluctuation of the economy during the 2007-2008 recession in the U.S | "Foreclosure", "Layoff", and "Recession" | Google search queries are worth pointing out the 2007-2008 peak turning point of the business cycle with both timeliness and accuracy |
| Baker & Fradkin (2017) | The impact of unemployment insurance policy on labor markets | Google Job Search Index (GJSI) | GJSI does react to the potential benefit duration of unemployment insurance, but the impact is small |
| Vosen & Schmidt (2011) | The private consumption (represented by consumer growth) | 56 weekly categories related to consumption from Google Trends | The prediction of Google factors outperforms the outcome of survey-based models (the University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index) |
| Balakrishnan & Dixit (2013) | The Future of the S&P500 Index traded at the Chicago Mercantile Exchange | 850 keywords that are strongly related to finance query the New York Times Article Search API | Google Trends data for financially relevant keywords can predict well the market volatility |
| Mccalum & Bury (2013) | The interests about the environment of the public | 19 series that relate to environment term to study | The strong negative slopes of search indicators during the period 2004-2009 |
| Martinez et al. (2021) | Grain prices | The queries related to three topics: Maize, Corn, and Soybean | Google Trends is valuable to predict grain prices |

## The application of Google Trends in Covid-19

There is a using Google Trends progress for predicting the activity and the recovery of economic growth in the Covid-19 pandemic (Keane & Neal, 2020; Fetzer et al., 2020; Eraslan & Gotz, 2020; Woloszko, 2020; Monache et al., 2021). A major amount of these papers consider that the contribution of Google Trends data is valuable in economic predictions. With the various categories and topics, Google Trends has the ability to cover the large of factors of the economy. Additionally, because the dataset covers countries that use Google search, it is convenient for the researcher to expand the scope of their research, or concentrate on their own countries. Furthermore, considered the most advantage of Google Trends, the multiple frequencies of dataset – from weekly frequency to monthly frequency – leads to timely reaction to the fluctuation of the economy during the crisis such as the Covid-19 pandemic. **Table 2.2 shows us several popular papers applying models including many predictors and some of them use the Google Trends dataset. The findings of the majority conclude that high-frequency data, especially Google Trends, improves the accuracy of the prediction because the reaction to the fluctuation during the Covid-19 pandemic is promoted. However, the challenge is identifying what elements having large impact on the economic activity as a result of the limitation of methodology to deal with large dataset. In the considerable affection for finding the way to interpret the relationship of prediction and predictors, Woloszko (2020) suggest SHAP tool is powerful to deal with it.**

Table . The need for big data in the Covid-19 pandemic research

|  |  |  |  |
| --- | --- | --- | --- |
| **Paper** | **Data** | **Method** | **Finding** |
| Eraslan & Gotz (2020) | - Weekly Google search queries related to "Unemployment", "Short-time-work", and "State support"  - Daily frequency indicators (Electricity, Toll, Pedestrian frequency, Consumer confidence, Air pollution, Flights) monthly frequency variables (IP). | Expectation-maximization algorithm | The tracking economic activity index performed quite well. Nevertheless, because WAI is not based on an econometric model, it is not able to compare directly with short-term predictions for GDP growth estimated by forecast models |
| Fetzer et al. (2020): | To research the development of economic anxiety, the authors leveraged Google search for topics ”Recession” and ”Stock Market Crash” for a total of 194 countries and territories.  To measure the capture panic reactions, search topics ”Survivalism” and ”Conspiracy Theory” are applied. | Descriptive Statistics | The results provided evidences that Google Trends performs well for showing the coranavirus pandemic leads to economic anxieties |
| Keane & Neal (2020) | Consumers’ concerns: ’panic buying’, ’panic’, ’hoarding’, and ’supermarket’  Product categories: ’toilet paper’, ’pasta’, ’rice’, ’flour’, ’vegetables’, and ’fruit’ | Ordinary Least Squares Regression | - The daily index of consumer built up by Google Trends data fit well with the timing of stockpilling as measured in the IRI data.  - The results show that the announcement of internal movement restrictions had effects largely vanished, although it just generated considerable consumer panic in the short-run.  - Stimulus announcements had a significant positive short-run impact on panic, but it was very short lived. Travel restrictions has no evidence showed effects on consumer panic.  - Both domestic and worldwide Covid-19 cases had impact on consumer panic. |
| Woloszko (2020) | Collected data search from Google of 215 categories and 33 topics that have major relevance to economic activity | Neural artificial network algorithm and SHAP tool | The output is used for tracking monthly and quarterly GDP in the past, and weekly GDP in the present. It performed well in showing the value of GDP. |
| Monache et al. (2021) | - Four "genuinely" weekly variables, named as gas pipelined to the industrial sector, consumption of electricity, trends in point-of-sale transactions (POS) from the payment system, and an index of volume searches on Google Trends of term "CIG" - Eight monthly variables, namely traffic flows on motorways, there expenditure indicators collected by ConfCommercio (goods, services, and their total), the purchasing managers' indexes (PMI) for both manufacturing and services, the total amount of CIG authorized work and the value-added tax on imports | Principal component model | The conducted index had a good predictive ability both out-of-sample during the Covid-19 pandemic and in-sample in "normal times". However, the relevance between ITWEI and real GDP quickly decreases over time. |

## Some concerns about Google Trends data

So far, the literature shows the usefulness of Google Trends data in economic research activity. Empirical papers provide the value of Google search indicators in the large area of fields of the economy, and this dataset is powerful for forecasting and nowcasting in fluctuation time likes the Covid-19 pandemic. However, Google search indicators are not only advantages but also disadvantage. Several papers point out some problems that the researchers are able to face when dealing the search queries. Firstly, the SVIs are generally the percentage value of the concerned keyword with the total queries searched in a particular period, it leads to the downward trend of several keywords over time because of the improvement of a variety of Google users (Stephens-Davidowitz & Varian, 2014), thereby some papers meet mistakes when concluding the interest of the public about particular topics then and now (Macclum & Bury, 2013; Ficetola, 2013). **Secondly, the relevance between Google Trends indicators and outcome variables does not maintain long, it requires updating frequently. Almost empirical works agreed that Google search indexes can react timely to the changes in fluctuation time, but several papers concerned that when the official indicators are released, the explanatory of Google search predictors declined rapidly (Woloszko, 2020); others supposed that if the estimate model having strongly correlated input variables,** Google Trends does not improve significantly forecasts’ accuracy (Combes & Bortoli, 2016); and even some works suggested do not compare the model based on Google Trends indicators to formal forecasting models (Eraslan & Gotz, 2020). Eventually, it is difficult to interpret the prediction models based on Google Trends predictors. Almost all papers that used Google search queries in the Covid-19 pandemic applied nonparametric estimates with many explanatory variables. It is not a suitable case for building a model with strong correlated independent variables (Giannone et al., 2018). Excepting the move of the economic activity at a particular time, detailed information of the impact of the factors can not be interpreted (Lewis et al., 2020).

**Google Trends is a powerful option for researchers, but it contains several pitfalls that require the researcher has to be careful. It shows that this dataset is obviously easy to reach but quite difficult to control.**

## Summary

Summarized, the Google Trends data shows the worth adopting in economic research activity. The fields applied this dataset is extremely broad: unemployment (Baker & Fradkin, 2017), private consumption (Vosen & Schmidt, 2011), recession (Chen et al., 2014), consumer confidence (Choi & Varian, 2011), finance (Balakrishnan & Kalpit Dixit, 2013), and even grain prices (Martinez et al., 2021), especially in the Covid-19 pandemic time (Keane & Neal, 2020; Fetzer et al., 2020; Eraslan & Gotz, 2020; Woloszko, 2020; Monache et al., 2021). **However, some papers concerned the accuracy of its estimators (**Stephens-Davidowitz & Varian, 2014; Macclum & Bury, 2013; Ficetola, 2013; **Woloszko, 2020;** Combes & Bortoli, 2016; Eraslan & Gotz, 2020; Giannone et al., 2018; Lewis et al., 2020), and previous works point out the challenge when dealing with big data is how to interpret the relationship between estimators and prediction, to find out what matter to the outcome during the study period.

# DATA AND METHODOLOGY

After reviewing the literature in the previous chapter, Chapter 3 will present data and methodology applied in this thesis. This chapter has three sections. The first is about the model, it provides detailed information on the two-step model. The second gives brief information about ANN algorithms and SHAP tool. The third section is about data, it provides the preprocessing data, and what Google Trends queries selected.

## Model

Due to the quick and deep effect of the Covid-19 pandemic on the economies, this thesis aims to construct a weekly tracking economic index to react timely to the impact of the pandemic. Many papers applying big data have been published in recent years, but the models applied were quite various. The most leveraged model is the two-step model, which is used by this thesis, but it is presented differently between research. This work decided to follow the two-step model presented by Woloszko (2020), which is suitable for the objective of this thesis. Aiming to construct an index tracking weekly economic activities, two-step model is a process that trains historical data with lower frequency in the first step and predicts weekly-frequency outcome from weekly-frequency predictors in the second step. The first step of model is as like a training step, which teaches how applying the machine learning algorithm on the dataset. This step uses training dataset, which includes quarter Google Trends and historical quarterly GDP growth rate.

|  |  |  |
| --- | --- | --- |
|  |  | (3.1*)* |

Where is a nonlinear function of the year-over-year log-difference of quarterly averages of search volume indices () for categories (indexed by 𝑐), and country dummies (), and is white noise. After the first model, the second model applies weekly Google Trends to construct Weekly Tracker:

|  |  |  |
| --- | --- | --- |
|  |  | (3.2) |

*WT* is the Weekly Tracker, which is interpreted as an estimate of the year-over-year growth rate of "weekly GDP" (same week compared to the past year).

## Estimation Methods

The large number of variables from Google search queries requires a method to deal with the large width dataset. The popular applied method is PCA (Principal component analysis), but previous papers point out this method has no ability to detail the relevance between predictors and outcome variables. This thesis uses ANN (Artificial Neural Network) algorithm to take prediction because of its potential to deal with the models having large number of variables. And after that, this work will apply SHAP tool to study which variables contributing considerably to the outcome of the prediction.

### Artificial Neural Networks algorithm

Motivated by the structure of the neurons of the human brain, ANN is a method that is considered valuable for forecasting in the case that having a large number of predictors (Elliot et al., 2016). Furthermore, ANN itself requires a large sample dataset, which is adopted well by Google Trends data, to perform the powerful value for prediction (Woloszko, 2020). The difference between PCA and dynamic factors of ANN is the non-linear approach (Elliot et al., 2016). It reduces the dimensionality to a number of intermediate components in the middle layer before making a prediction (Woloszko, 2020). With the multi-layer structure, ANN can avoid overfitting. It has an ability to deal with the extremely large sample with high dimensions, even the number of predictors is higher than the number of the observations (Csaji, 2001). Therefore, this algorithm has been increasingly applied in academic research in recent years, especially with the rapid improvement of media (Elliot et al., 2016; Aiken, 2000; Jena et al., 2020; Malte, 2018; Loerman & Benedikt, 2019).

A neural network is fundamentally operated by an information-processing called as a *neuron*. The model of a neuron (basic form of an artificial neural networks) has five main elements: *input signals*, which are predictors of this thesis, are maybe plural; *synapses or connecting links*, which is variables value, show the strength of the connection of each input signal to a neuron; an *adder* takes a weighted sum of input signals; an *activation function* for optimizing the weights, or by other works are limiting the neuron output; an *external bias* is included in the model to increase or decrease the net input of the activation function; and *output*, which is outcome variable of this thesis. The model of a neuron is delineated by figure 3.1.

**.**

**.**

**.**

Input signals

X1

X2

Xn

**.**

**.**

**.**

Synaptic weights

Summing junction

Activation function

Y

Output

φ(.)

Bias (b)

Figure . The model of a neuron (Csaji, 2001)

The model of a neuron can be described in mathematical term by:

|  |  |  |
| --- | --- | --- |
|  |  | *(3.3)* |

Where is input signals, are the synaptic weights of the neuron, *b* is the bias, is the activation function and is the output of the neuron. Following Woloszko (2020), the model of this thesis uses "relu" activation function.

The neurons in a layered neural network are organizational layer form. At least, a layered neural network has two layers: an *input* layer and an *output* layer. Layers which are between input layer and output layer called as *hidden layers*, whose have *hidden neurons* or *hidden units* that are computations nodes. The neural network having the input of the neurons of each layer is the output of the preceding layer only, not vice versa, called *multilayer feedforward architecture* (figure 3.2). As said by Csaji (2001), by adding more hidden layers, high-order statistics may be extracted by computing the network. By other words, in a rather loose sense, although it is potentially *local* connectivity due the extra synaptic connections and the extra dimension of neural interactions, the network is able acquires a *global* perspective. The neural network of this thesis uses two hidden layers of 300 and 10 neurons.

Figure .: Multilayer feedforward architecture (Csaji, 2001)

The main concern of this method is their black-box nature, which is addressed by machine learning interpretability techniques presented below.

### SHAP tool

SHAP (SHapley Additive exPlanations) is an interpretability tool used for open the black box[[4]](#footnote-4). It is released by Lundberg and Lee (2016) to explain the prediction of each instance due computing the contribution of each feature (for this thesis it is input variable) to the prediction. SHAP is constructed by the concept of the Shapley values, a method from coalitional game theory that is built to fairly distribute a "pay-out" from a multi-player game (Woloszko, 2020; Molnar, 2020). In this matter, the "game" is a single instance prediction task, the "players" are the predictors, and the "pay-out" equals actual this instance prediction minus the average prediction for all instances (Molnar, 2020). The Shapley value is the average marginal contribution of an estimator value to the prediction across all possible coalitions, which is designed as a number of predictors taking the value that is noticed rather than their average or any arbitrary value. The model prediction is added up by Shapley values. In the mathematical term, for simple linear model prediction described by:

|  |  |  |
| --- | --- | --- |
|  |  | *(3.4)* |

Where is the instance which is desired to compute the contributions. is a feature value, with j = 1,…,p. The is the weight corresponding to predictor j. The contribution of the j-th feature on the prediction is:

|  |  |  |
| --- | --- | --- |
|  |  | *(3.5)* |

Where represents the mean impact estimate for feature j. The contribution equals the feature impact minus the average impact. The following is the result when taking sum all the feature contributions for one instance:

|  |  |  |
| --- | --- | --- |
|  |  | *(3.6)* |

|  |  |  |
| --- | --- | --- |
|  |  | *(3.7)* |

As Woloszko (2020) said, Shapley value is the only ascription technique that associate the following properties: efficiency (Shapley values sum to the prediction minus its average), symmetry (variables which have the Shapley value should have equal contribution to all coalition), dummy (a variable value with no impact on the prediction whatever the coalition has Shapley value equal to zero) and additivity. The considerable value of this method is the tendency to provide both local interpretability (explaining given predictions) and global interpretability (the general functioning of a model).

## Data

This thesis tracks the economic activity of 22 countries in Asia and Pacific, the period for training data is from the first quarter of 2004 to the second quarter of 2021, and then a weekly economic index is constructed for 2021. The training dataset is panel data with around 1400 observations. This thesis decides to pool countries together to widen the sample size and thus estimate precision. The large sample of cross-country may lead to the heterogeneity, but as long as country dummies are included, this problem is able be handled by multilayer feedforward architecture. Each probable interaction between Google Trends estimators and country dummies can be flexibly modeled by multiple hidden layers with enough neurons. The input signals of each neuron are Google Trends predictors, country dummies and monthly variables.

### Outcome variable

This thesis chooses the dependent variable is GDP growth rate, which is the most popular index used for reflecting the economic activity (Monache et al., 2021; Jardet & Meunier, 2020; Lewis et al., 2020; Woloszko, 2020; Tamara et al., 2020; Zhemkov, 2021). The data is collected from the database of ADB, with quarterly frequency (year-over-year difference as same period).

### Google Trends variables

This thesis applies time-series option Google Trends. The search volume index is concluded from below formula:

|  |  |
| --- | --- |
|  | (3.8) |

Where is the search volume index of query *c* in time *t*, is search volume of query *c* in time *t*, and is the search volume of total queries in time *t*. The search volume indices are the percentage ratio and they are affected by time, it is needed for preprocessing data for search series variables before applying to estimate. Woloszko (2020) suggested that take logarithm-term to reduce the trend effect.

|  |  |
| --- | --- |
|  | 3.9 |

Furthermore, Google Trends data provides vast variables, although it leads to the availability to cover a large area of economic components, it is demanded that to select carefully necessary categories and topics because of the limit of the estimation method.

Previous works proved that the categories queries and topics queries are more valuable for applying in academic research than specific keywords (Woloszko, 2020; Choi & Varian, 2011; Suhoy, 2009; Martinez et al., 2021). Therefore, this research puts in the prediction 80 topics and 100 categories that are related to economic factors. All queries are collected with monthly frequency for the training step, and then weekly frequency for the prediction step.

The topics and categories are grouped by nine terms: Google Trends data is grouped by: Economic Anxiety, Health Condition, Consumption Services, Consumption Goods, Labor Market, Housing Construction, Finance, Business Services, Industrial. Because literature points out that the Google Trends dataset is affected by time trend, natural log of SVIs are take to avoid this problem. Furthermore, to address the seasonality, categories-based predictors are transformed using the log difference with the same period of the past year. Meanwhile, because topics-based variables are less impacted by seasonality, they are not different. The detail of the Google search queries, monthly predictors and output variable of this thesis will be presented in Appendix A.

This thesis uses two dataset Google Trends. First dataset includes data collected in period from 2005-2021 for training step. Because the data is reported in monthly-frequency, it is take quarter average to add in the model to apply algorithm for training with dependent variable. The second dataset includes data collected in last three years. It already has been taken weekly and this thesis uses it to estimate the Weekly Tracker.

# RESULTS

This chapter has three main parts.

Section 4.1 will focus on the performance of the outcome from first step of the model. The outcome of ANN algorithm is provided and proved that outperforms the prediction of AR(4) model after comparing the RSMEs. In the addition, section 4.1 shows the contribution of variables in historical dataset to the output of model through Shapley values.

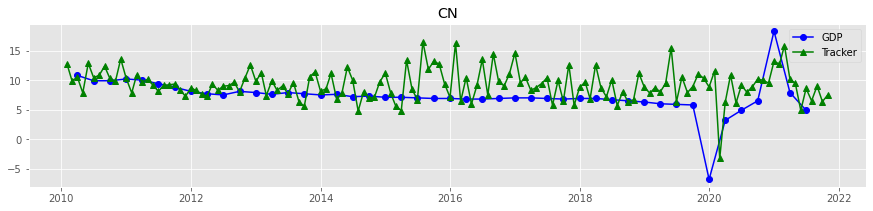
Section 4.2 concentrates on the Weekly Tracker. It will show how this tracker fluctuates during 2021, especially in the third quarter, to provide a view of the impact of the new Covid-19 wave in Asia and The Pacific countries. Furthermore, Shapley values are provided to point out which factors having important roles for the economic growth.

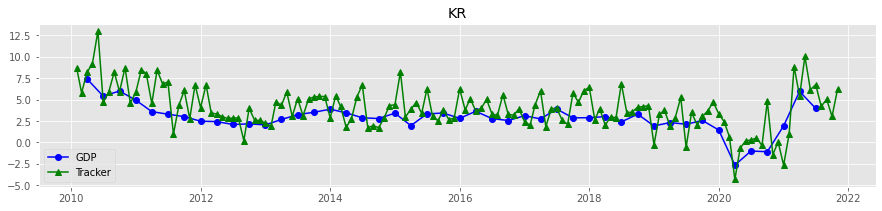
Section 4.3 is discussion the result of the model. The detail analyzes the trend of the growth rate in 2021 Q3 and the first haft of Q4 are provided. Furthermore, the case of Vietnam will be mentioned in this section.

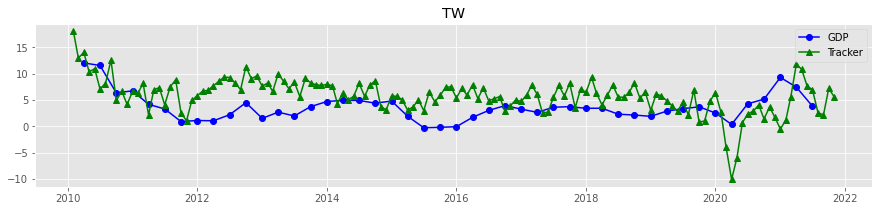
## Outcome from ANN algorithm

### Performance of the output from the training step

Pseudo-real time simulations on the quarterly series is applied to assess the predictive performance of the model. By looping on time and employing only past or present data, pseudo-real time simulations imitate the conditions a forecaster would have faced at each time period (Woloszko, 2020). The algorithm is trained on Google Trends and GDP growth rate data up to the first quarter of 2021, while the simulations will perform up to November of 2021. The simulations provide evidences that the outcome from model supplies relevant dominant information on GDP growth, economic crises and business cycles in almost all 22 Asia and The Pacific economies in the sample. The simulations suggest a slightly flop of economic performance in the third quarter and first two months in the fourth quarter of 2021.







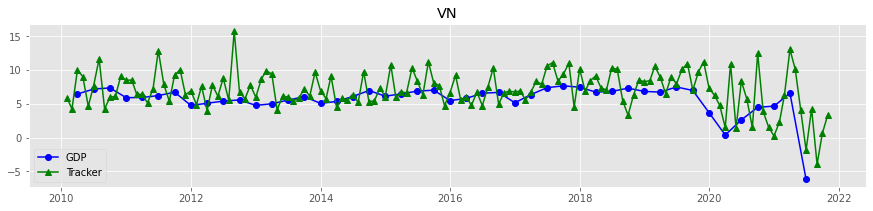


Figure . Outcome from model and actually quarter GDP growth rate (year-over-year).  
Source: Calculated by the author.  
Note: Figure show the result of four countries having reported the third quarter data, other countries will be provided in the Appendix B.

Table 4.1 provides evidence that the tracker is valuable to perform the GDP growth rate with past data. For the period from 2005 to 2021, the Root Mean Squared Error (RMSE) of the Google Trends and monthly variables tracker over the autoregressive (AR) model has a median is 0.43. The period 2008 -2011 is used for representing the fluctuation time (the Global Financial Crisis and Debt Crisis in Europe), and we can see that the RMSE of the tracker is equal to 0.36 AR model. With the period that we concentrate, the performance of tracker even has the improvement with RMSE equaling just 0.33 AR model. These are pieces of evidence to consider that the error of the outcome from ANN algorithm is smaller than the error of the result from AR(4) model. It is demonstrated that the outcome of the model is powerful to reflect GDP growth rate of 22 economies in this thesis sample.

Table . Forecast performance, comparing to AR(4) model.  
Notes: The first column (Unadjusted) signifies root mean square error in forecast simulations. RMSEs are standardised by the GDP growth rate standard deviation in the second column (Standardised), and divided by the RMSE of an AR(4) model in the third one [Relative to AR(4)]. The period of the full sample is from 2005 Q1 to 2021 Q2, Crisis is from 2008 Q1 to 2010 Q4, and the Covid-19 is from 2020 Q1 to 2021 Q2.  
Source: Calculated by the author

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2005-2021 (Full sample)** | | | **2008-2011 (Crisis period)** | | | **2020-2021 (Covid-19 pandemic period)** | | |
|  | **Unadjusted** | **Standardised** | **Relative to AR(4)** | **Unadjusted** | **Standardised** | **Relative to AR(4)** | **Unadjusted** | **Standardised** | **Relative to AR(4)** |
| Armenia | 1.75 | 0.23 | 0.31 | 2.77 | 0.29 | 0.27 | 2.31 | 0.26 | 0.27 |
| Australia | 0.74 | 0.39 | 0.56 | 0.78 | 1.77 | 1.43 | 1.62 | 0.32 | 0.43 |
| Azerbaijan | 2.99 | 0.29 | 0.80 | 2.14 | 0.55 | 0.46 | 1.38 | 0.32 | 0.32 |
| Brunei Darussalam | 1.94 | 0.46 | 0.52 | 0.69 | 0.19 | 0.23 | 3.02 | 1.50 | 2.10 |
| China | 1.04 | 0.31 | 0.41 | 0.80 | 0.42 | 0.47 | 2.28 | 0.33 | 0.33 |
| Georgia | 1.94 | 0.30 | 0.41 | 0.69 | 0.08 | 0.28 | 5.60 | 0.40 | 0.49 |
| Hong Kong | 1.00 | 0.25 | 0.50 | 0.96 | 0.18 | 0.36 | 1.85 | 0.26 | 0.46 |
| Indonesia | 0.72 | 0.31 | 0.55 | 0.94 | 0.89 | 0.89 | 1.23 | 0.29 | 0.33 |
| India | 1.75 | 0.32 | 0.41 | 0.63 | 0.17 | 0.27 | 5.58 | 0.39 | 0.42 |
| Japan | 0.64 | 0.22 | 0.33 | 0.99 | 0.20 | 0.36 | 0.81 | 0.15 | 0.19 |
| Kyrgyz Republic | 2.90 | 0.29 | 0.39 | 5.40 | 0.36 | 0.41 | 3.82 | 0.49 | 0.38 |
| Korea | 0.56 | 0.27 | 0.45 | 0.53 | 0.14 | 0.28 | 0.50 | 0.18 | 0.21 |
| Kazakhstan | 1.07 | 0.29 | 0.46 | 0.94 | 0.20 | 0.24 | 1.45 | 0.48 | 0.34 |
| Sri Lanka | 2.49 | 0.43 | 0.52 | 3.87 | 0.39 | 0.49 | 3.36 | 0.39 | 0.39 |
| Mongolia | 1.91 | 0.28 | 0.33 | 1.84 | 0.33 | 0.29 | 2.69 | 0.35 | 0.31 |
| Malaysia | 0.90 | 0.21 | 0.30 | 1.03 | 0.19 | 0.32 | 1.99 | 0.20 | 0.24 |
| New Zealand | 0.85 | 0.29 | 0.39 | 0.47 | 0.22 | 0.22 | 1.69 | 0.22 | 0.30 |
| Philippines | 0.97 | 0.21 | 0.38 | 1.03 | 0.36 | 0.50 | 1.36 | 0.15 | 0.19 |
| Singapore | 2.89 | 0.56 | 0.95 | 6.22 | 0.72 | 1.61 | 2.40 | 0.28 | 0.34 |
| Taiwan | 2.44 | 0.65 | 1.33 | 4.92 | 0.61 | 2.16 | 2.18 | 0.79 | 0.91 |
| Thailand | 1.26 | 0.31 | 0.45 | 0.78 | 0.15 | 0.26 | 1.18 | 0.20 | 0.22 |
| Viet Nam | 0.77 | 0.34 | 0.39 | 1.03 | 0.35 | 4.30 | 1.19 | 0.30 | 0.26 |
| Median | 1.16 | 0.30 | 0.43 | 0.98 | 0.31 | 0.36 | 1.92 | 0.31 | 0.33 |

Figure 4.2 focuses on the time of the new Covid-19 wave. Circles are official report GDP growth rate, and triangles represent for tracker index. We can see that it is not quite different between the position of circles and triangles. Even some countries have circles and triangles being in the same position, such as Australia, Azerbaijan, Hong Kong, Indonesia, Korea, Malaysia, Taiwan, and Vietnam in 2021 Q2, and Korea and Taiwan in 2021 Q3. However, this figure demonstrates the concern about Covid-19 shock has ability to challenge the predict based on the historical data.

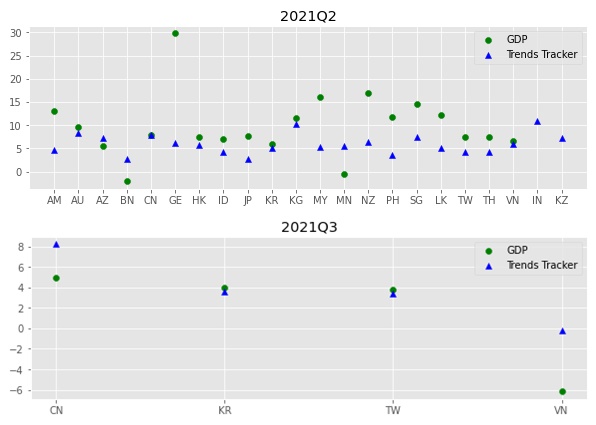


Figure . Tracker's prediction for 2021 Q2 and Q3  
Notes: With the 2021 Q2, figure show all countries in sample except India and Kazakhstan. While 2021 Q3, only focus on four countries that reported GDP growth rate of the third quarter.  
Source: Calculated by the author  
Data source: ADB

Summarized, this section provides evidences to prove that the outcome from ANN algorithm is reliable. The simulations show the relevant reflection for GDP growth rate of historical data. Comparing RSMEs demonstrates that the Trends Tracker is more reliable than AR(4) model output. And the predictions for 2021 Q2 and Q3 capture brilliantly the reaction of GDP growth rate to the shock of Covid-19 new wave.

### The contribution of variables to the outcome

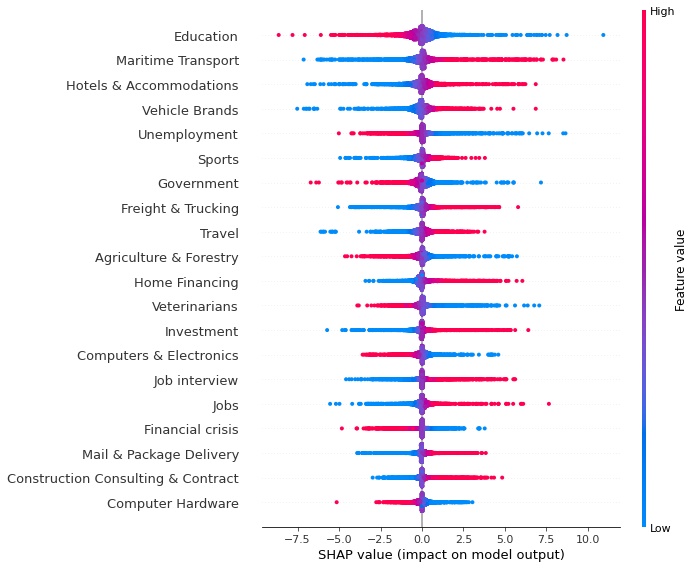


Figure . Most important variables and their contributions to predictions  
Note: Shapley values are the contributions of variables to the GDP growth rate estimated by the model. Ranking variables bases on importance, and for each variable. Each point indicates to an observation (that is a given month \* a given country) and its color depends on the value of the variable.  
Source: Calculated by the author

Figure 4.3 provides variables having large contributions to the GDP growth rate estimated by the model. Query related to "Education" has the most contribution to the predictions, however, the relationship is negative. When searches of "Education" increases, the outcome decreases. The second important variable is "Maritime Transport", as expectation about of the contribution of transport to GDP growth rate estimate, the impact on model output of this query is positive. Queries related to Consumption ("Hotel & Accommodations", "Vehicle Brands", "Sports", " Freight & Trucking", "Travel", "Mail & Package Delivery") have large donation for the growth rate of 22 countries in the sample. While queries related to Economic Anxiety ("Unemployment", "Government", "Financial Crisis") have significantly negative effects on the output of model. Furthermore, queries related to Finance ("Home Financing", "Investment"), Labor Market ("Jobs", "Job Interview"), Hom Construction ("Construction Consultant & Contract") contribute an important positive impact on the outcome.

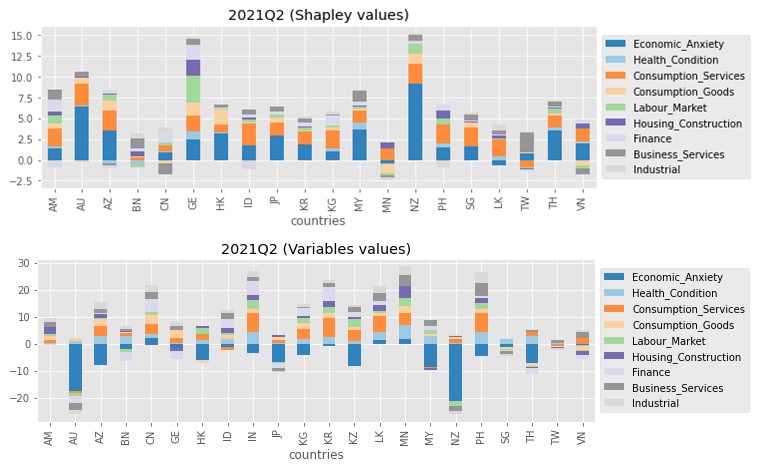


Figure . A focus on 2021 Q2  
Note: Google Trends variables are aggregated together into groups described in Appendix A.  
Source: Calculated by the author

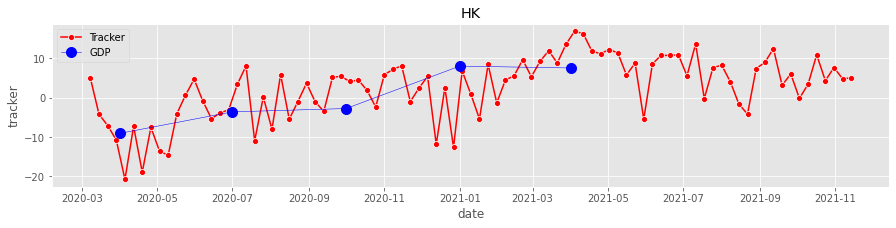
Figure 4.4 provides insights on the predictions for the second quarter of 2021. It shows the important variables which have large contribution across countries. For the sake of readability, this thesis aggregates Google Trends variables together into significant groups. In 2021 Q2, economies still be in the recovery cause of the decrease of the Covid-19 shock in 2020, and the estimate GDP growth rate reflect this situation. Interestingly, Economic Anxiety seems to be the most important factor of the recovery. This contribution can be explained by the decrease of searches related to Economic Anxiety, which is described by variable values. The second largest contribution belongs to Consumption Services, while Consumption Goods has not considerably impact on the output. Labor Market is another one playing key role for the output of the model. Although Health Condition topics and categories are popular on Google in 22 countries in sample, its contribution to outcome is not significant.

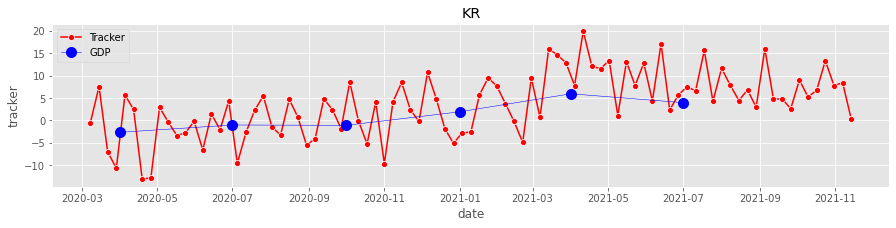
## Weekly analysis

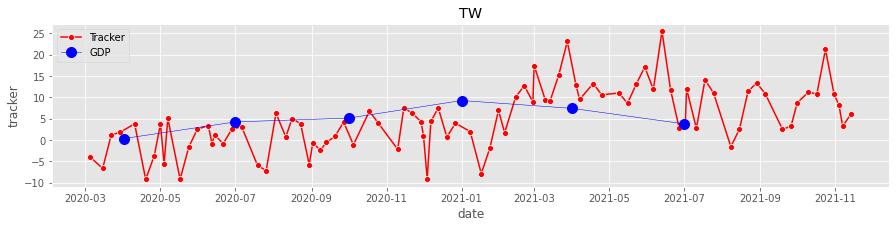
### The fluctuation of Weekly Tracker during the new wave of Covid-19

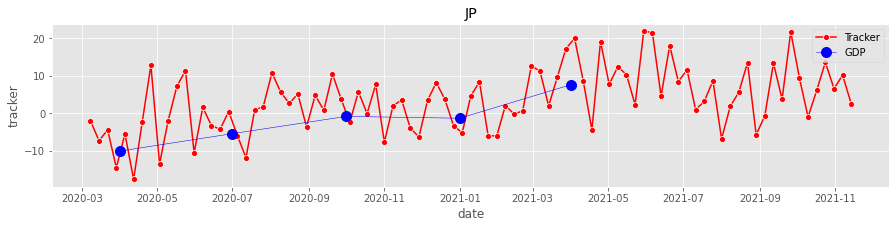
The model of GDP growth rate is presented in Chapter 3 was trained using data at a quarterly frequency, and is being applied to produce a Weekly Tracker in this section. The Weekly Tracker provides weekly information on business by the time of the writing. This section concentrates on the period during the new wave of the Covid-19 pandemic. The Weekly Tracker provides valuable insights on the GDP growth rate during the Covid-19 pandemic. Figure 4.5 and 4.6 shows that how the impact of the new wave of Covid-19 constrains the economies, even developed countries. In the first haft of 2021, almost all countries in the sample find out the recovery after the epidemic shock in 2020, but it is cut soon in the third quarter of 2021, and the stagnation of growth rate may maintain to the end of 2021.

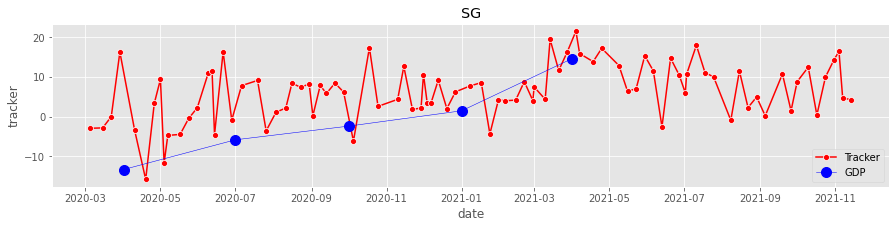
In the second quarter of 2021, advanced countries get a quick recovery. According to ADB, the GDP growth rate of Hong Kong is 7.5%, Korea is 5.9%, Taiwan is 7.4%, Japan is 7.6%, Singapore is 14.6%, Australia is 9.6% and New Zealand is 17%. Numeracies show the brilliant efforts of these countries when compared to the flop during 2020 Q2 and Q3. However, Weekly Tracker points out the concern about the significant negative impact of the new wave of Covid-19. This shock reduces the weekly GDP growth rate estimate close to 0% and even negative at some points. Figure 4.5 provides insights on the stagnation of GDP growth rate when the Weekly Tracker fluctuates around 0%. Hong Kong has a negative Weekly Tracker in June and August, it is similar to Taiwan, Japan, and Singapore. The Tracker of these countries sometimes recovers quickly to 10% or even 20%, but the trend is horizontal and this situation has no signal to stop in the fourth quarter of 2021. Korea and New Zealand are two cases that the Weekly Tracker that has not been smaller than 0%, but the trend in the 2021 Q4 shows that the expectation about the flop of GDP growth rate estimate. Australia is the only one having the signal of recovery when the trend of Weekly Tracker is upward from August to November.

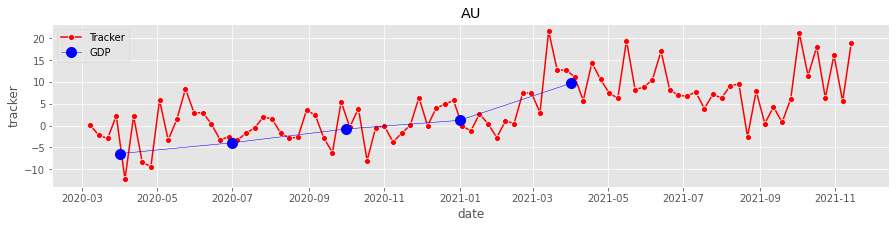












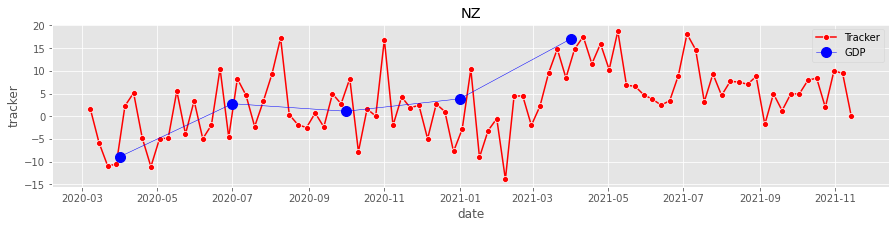
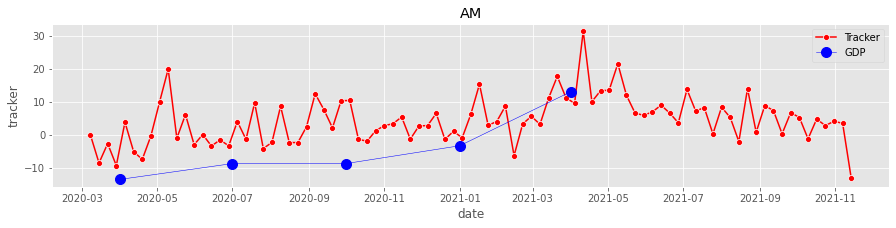
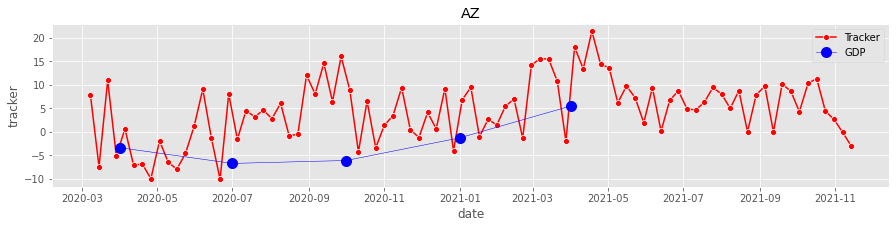
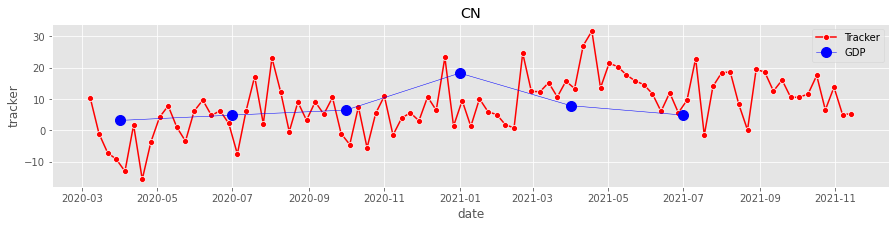


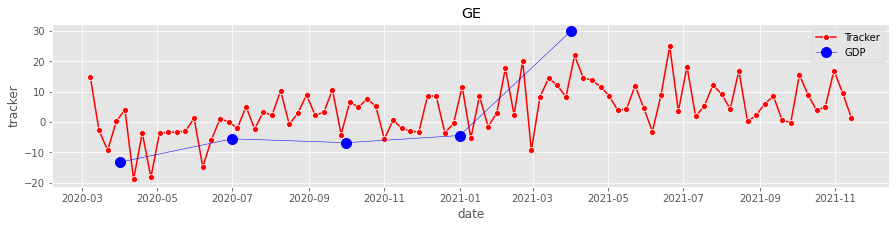
Figure . Weekly Tracker: High-income economies (According to World Bank)  
Source: Calculated by the author

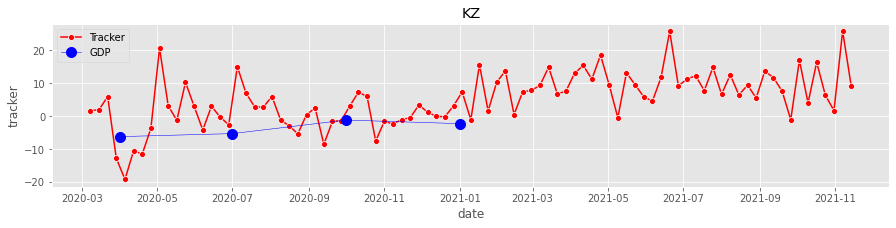
Comparing to high-income economies, the situation of upper-middle countries seems to be more concerns. During the second quarter of 2021, same as advanced countries, Armenia (13.3%), Azerbaijan (5.4%), China (7.9%), and Georgia (29,8%) get a high GDP growth rate. However, a steep down trend of Weekly Tracker has been immediately recorded since June. Officially, the GDP growth rate of China is reported lower in the third quarter, 4.9%. Meanwhile, the Weekly Tracker of Kazakhstan is slightly upward during 2021 Q2 and Q3, even to October and November. However, the GDP growth rate of Kazakhstan did not recover significantly during 2020 Q4 and 2021 Q1, when the official reported number was still smaller than 0%, it means this country faced arduousness cause of its low GDP during the pandemic. The Tracker trend of Malaysia and Thailand during 2021 Q3 and Q4 seems to be the same as high-income economies, it is unstable but horizontal. Although Malaysia is slightly upward, it strongly fluctuates, such as the lowest was close to -10% while the highest was close to 30%. Tracker trend of Thailand is more consistent than Malaysia but it pretty downgrades.

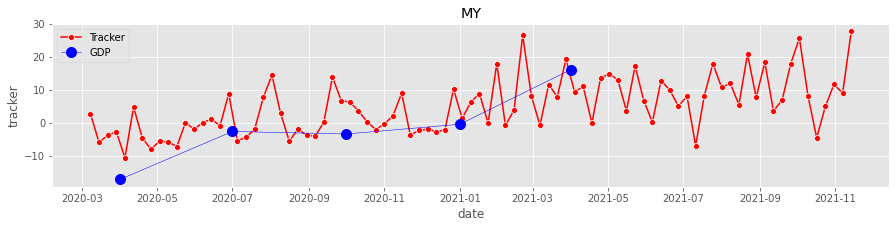












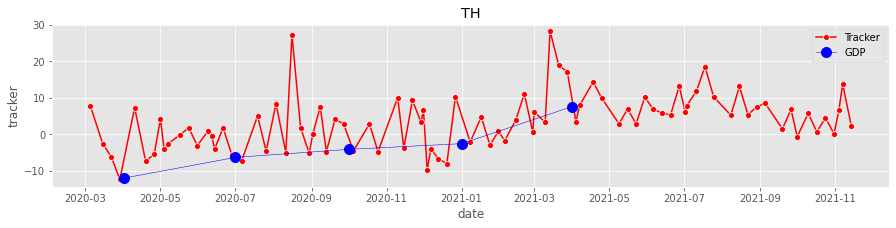


Figure . Weekly Tracker: Upper-middle income economies (According to World Bank)  
Source: Calculated by the author

Compare to advanced countries and upper-middle-income countries, the Weekly Tracker of lower-middle-income economies appears more stable during 2021 Q3 and Q4. In the second quarter of 2021, same as advanced countries and upper-middle-income economies, the lower-middle-income group has a significant recovery. According to ADB, the GDP growth rate of Indonesia is 7%, Kyrgyz Republic is 11.5%, Philippines is 11.8%, Sri Lanka is 12.3% and Vietnam is 6.5%. The impact of the new wave of Covid-19 spiked the upward trend of GDP growth rate estimate in the first haft of 2021, however, the movement of Weekly Tracker is considerably consistent during 2021 Q3 and Q4. Indonesia, Philippines, and Sri Lanka remained the positive value of Weekly Tracker, while Mongolia and India had the stable motion during the first haft of 2021 Q4. Vietnam has the consistent trend of the Weekly Tracker and its value keeps positive in the time of 2021 Q3 and Q4, but the official reported GDP growth rate of 2021 Q3 is -6.17%, which is strongly lower than the Tracker. It points out the consideration that the impact of the Covid-19 shock challenges the reaction of historical data.

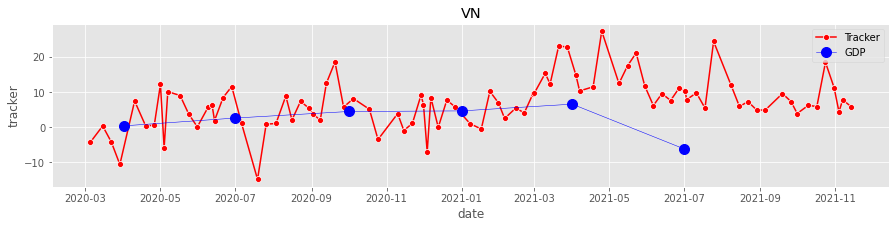
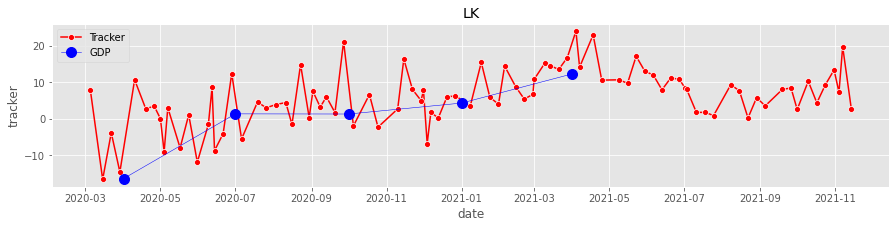
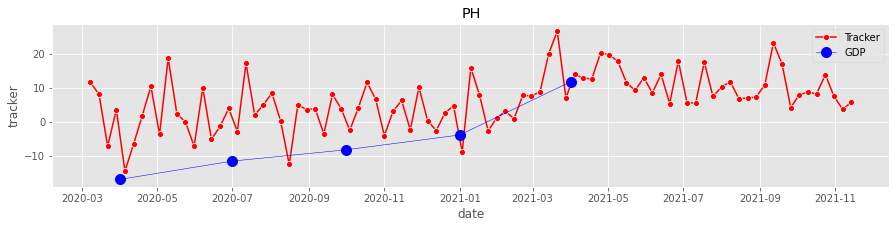
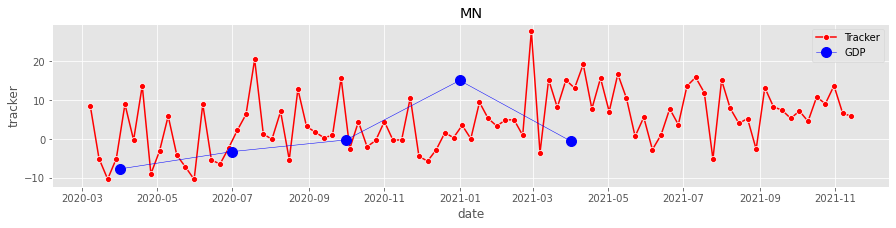
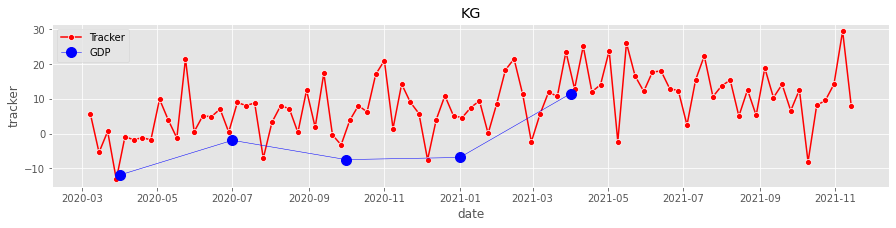
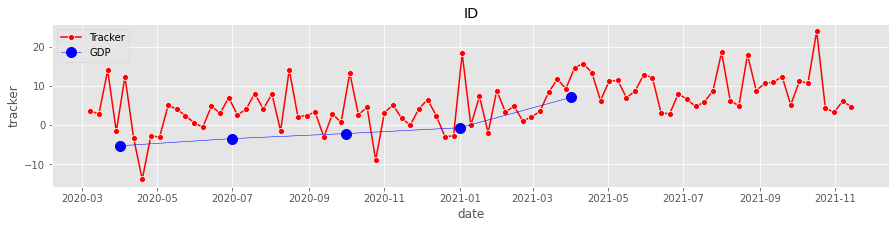
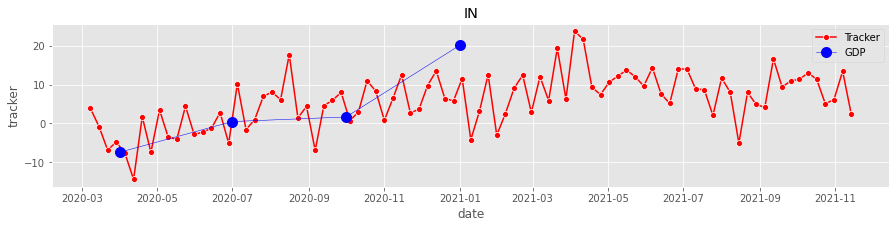


Figure . Weekly Tracker: Upper-middle income economies (According to World Bank)  
Source: Calculated by the author

### The positive signal at the end of 2021

Although the impact of the Covid-19 shock cut the recovery movement of almost all countries in the sample of this thesis, the positive signal appeared in the fourth quarter. Figure 4.8 focuses on the most recent predictions of GDP growth rate. Panel A shows the Weekly Tracker of mid-November is higher than the average Weekly Tracker of the period from July to November of 2021 (excepts Australia, Brunei, and Malaysia), it points out the chance to get over the stagnation cause of the pandemic. Furthermore, panel B provides more shreds of evidence that the growth rate estimates move positively. According to the report in September 2021, ADB reduces the GDP growth rate forecast 2021 of developing Asia from 7.3% to 7.1%. Almost economies from South Asia and Southeast Asia is decreased the growth rate estimations by ADB, such as Sri Lanka (4.1% to 3.4%), Brunei (2.5% to 1.8%), Indonesia (4.5% to 3.5%), Malaysia (6% to 4.7%), Thailand (3% to 0.8%) and Vietnam (6.7% to 3.8%). The new wave of Covid-19 may be the reason that leads to the concerns of economists. There is no evidence that the pandemic stops and it has no impact on the economies, but through the Weekly Tracker, it is found the expectation that the economies have tendencies to recover at the end of 2021 if the Covid-19 can be constrained because of the much higher of values.

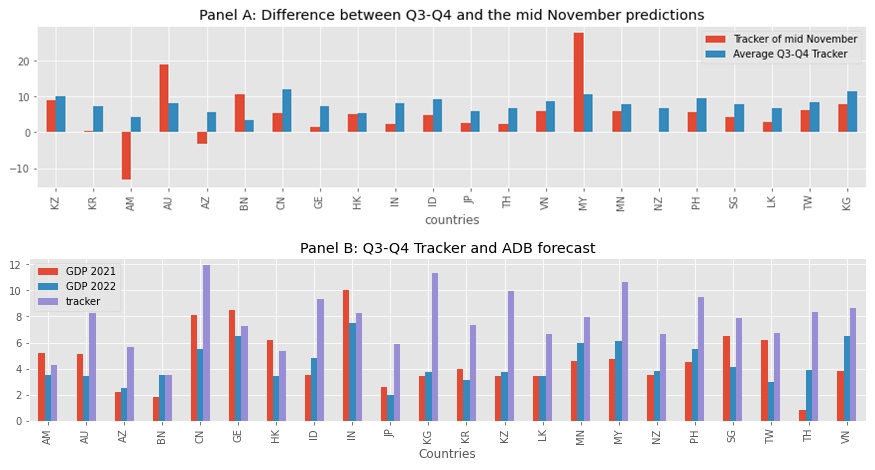


Figure . Most recent predictions of the Weekly Tracker  
Note: "Q3-Q4 Tracker" of both panels is the average value of the Weekly Trackers during Q3 and the first haft Q4 of 2021. In panel B, the GDP growth rate of 2021 and 2022 is the forecast of ADB.  
Source: Calculated by the author  
Data: ADB

### The factors having large contribution to the growth rate during the Covid-19 pandemic

Figure 4.9 shows us the contribution of variables to the Weekly Tracker estimates since the Covid-19 appearance. Applying Shapley values, this thesis finds out the key role of transport for economic recovery, when Google searches related to "Maritime Transport" have the largest contribution to the GDP growth rate estimates. In addition to, topics and categories related to Consumption (for examples: "Hotels & Accomodations", "Travel", "Sports", "Freight & Trucking", "Travel Agencies & Services", "Vehicle Brands", "Baggage"), Labor Market ("Job interview", "Jobs", "Civil service entrance examination"), Finance ("Investment", "Return of investment") and Housing Construction ("Construction Consulting & Contract) also contribute positively to the Weekly Tracker. By contrast, topics and categories related to Economic Anxiety ("Unemployment", "Financial crisis", "Government") harm the predictions. Interestingly, queries related to "Education" and "Veterinarians" contribute strongly negatively to the Weekly Tracker.

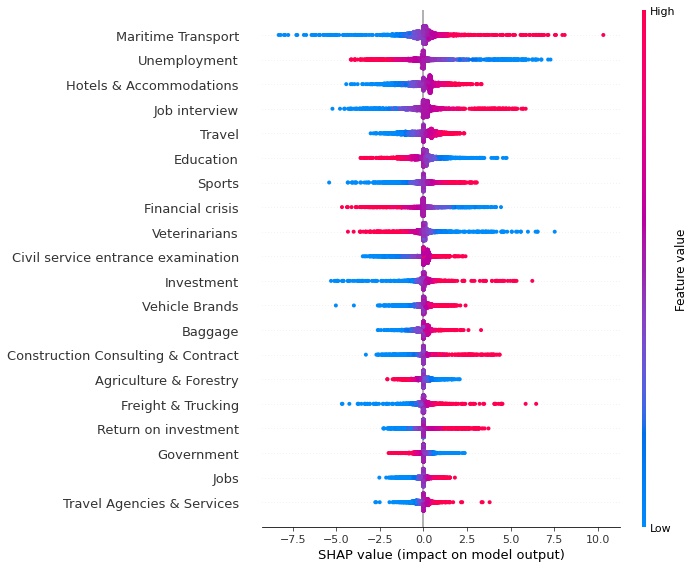


Figure . Contribution of variables to predictions during the Covid-19 pandemic  
Source: Calculated by the author

Figure 4.10 gives more detail of the contributions of variables to the GDP growth rate estimate. Consumption Services, Consumption Goods, Labor Market and Finance help significantly for the recovery of the Weekly Tracker. During the 2021 Q3 and the first haft of Q4, topics, and categories related to Consumption Services considerably increase and they impact positively on the Weekly Tracker. Meanwhile, despite the small growth, Consumption Goods still effects positively model output. The growth of topics and categories related to the Finance and Labor Market in July, September, and October brings to the Weekly Tracker a positive impact. However, because of the decrease during August, the relationship between Finance, Labor Market, and the predictions is not clear. Business Services topics and categories also help the recovery of the Weekly Tracker, especially in November. By contrast, Economic Anxiety topics and categories have an extreme negative relation with the Weekly Tracker. The downgrade of them during 2021 Q3 and the first haft of Q4 push significantly the growth rate estimates. On the other hand, although search number Heal Condition topics and categories improved brilliantly, their contribution to the output of the model is not considerable.

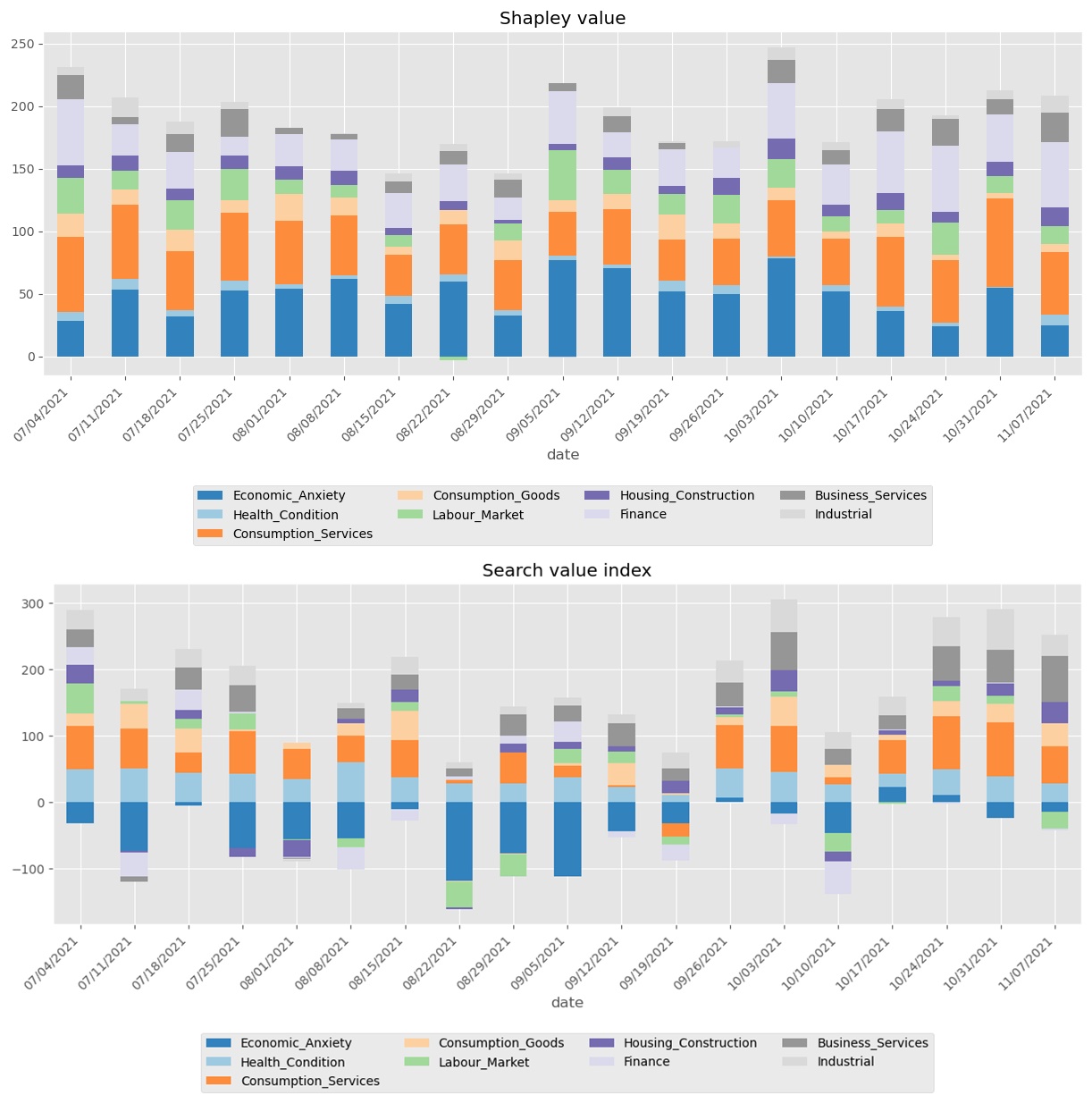


Figure . Drivers of the growth rate during the Covid-19 new wave   
Note: Contributions of variables to outcome are represented by Shapley Value. The search value index is indicated by the difference of svi year-over-year. Variable contributions and year-over-year differences svi are aggregated into economically relevant subgroups reflecting key economic sectors, after taking the sum of all countries in the sample.  
Source: Calculated by the author

## Discussion

### The expectation of the recovery at the end of 2021

The Weekly Tracker suggests the upward trend of GDP growth rate at the beginning of 2021 has been cut off since the new wave of the Covid-19 and the economic growth rate is constrained during the second haft of 2021 for 22 Asia and The Pacific countries in this thesis sample. The positive point is the impact of the Covid-19 during 2021 Q3 and Q4 seemly does not make a downward movement for the economic growth rate, as Weekly Tracker points out. However, the pandemic needs to be kept in control for guaranteeing that economies not to be hurt because of the impact of the fear, the reduction of consumption, and the risk closing trade.

Table 4.2 provides categories and topics which are the most searched since the beginning of the Covid-19 pandemic. Economic Anxiety topics and categories are extremely popular during the coronavirus spread out, especially in the beginning. The fear of people during the pandemic is performed by the large contribution of topics and categories related to recession and bankruptcy. The Covid-19 impacted the economy, it leads to people worry the economic situation, both micro and macro. In the case of microeconomic, individuals concern the situation of companies, factories, or small private businesses, so they search for information about bankruptcy to study the situation and supports from the government. And for macroeconomic, people can relate this pandemic to previous recessions such as the financial collapse in 1997, 2008 to predict or estimate the future. Because of the fear, people are also increasingly concerned about their financial issues. They worry about their loan, interest rate, or government subsidy. It leads to the results that the number of queries related to unemployment and personal finance is large and contributes significant negation to the economic growth rate. The disappearance of topics and categories related to Economic Anxiety in table 4.2 points out the reduction of fear of people. This reduction should remain because of the relation with the Weekly Tracker shown in Figure 4.10.

The growth of the number of topics and categories related to Consumption points out the adaptation to the pandemic situation. Table 4.2 shows that the Consumption variables have been listed since 2020 Q4, the time that almost all economies found the recovery. During the first shock of the Covid-19, many countries applied travel limitation protocols, it makes people have to stay at home for a long time and to be familiar with shopping online and entertainment at home. Moreover, when the fear about the pandemic decreases and the travel limitation protocols are reduced in almost all countries, the demand constrained during lockdown or social-distancing time pushed the Consumption searches on Google since the first quarter of 2021. It leads to the growth rate of search value index of topics and categories related to Consumption during 2021 Q3 and the first haft of Q4 (as presented by Figure 4.10). In the addition, trade has an important role in the GDP growth rate. The queries related to "Maritime Transport" have the most positive impact on the Weekly Tracker, according to the Shapley value. The appearance of topics and categories "Maritime Transport", "Distribution & Logistics" can be considered the motivation for keeping the Weekly Tracker still be positive during the new wave of the Covid-19.

Although the contribution to the Weekly Tracker is not quite significant, Heal Condition topics and categories have impressive growth during 2021. Table 4.2 shows that 'Pharmaceuticals & Biotech' have always been the top topic in both all groups of income economies. In addition, "Doctor's Offices" and "Hospitals & Treatment Centers" receive a large interest. It provides insights on the concern and knowledge about the pandemic of people.

Table . Top categories and topics since the coronavirus spread out  
Source: Calculated by the author

|  |  |  |  |
| --- | --- | --- | --- |
|  | **High Income Economies** | **Upper Middle Income Economies** | **Lower Middle Income Economies** |
| 2020Q1 | 'Stock market crash', 'Fixed-rate mortgage', 'economic crisis', 'Financial crisis of 2007–2008', 'Recession' | 'Stock market crash', 'Recession', 'economic crisis', 'Coronavirus recession', 'Financial crisis of 2007–2008' | Recession', 'Stock market crash', 'Global recession', 'Financial crisis', 'Coronavirus recession' |
| 2020Q2 | Recession', 'Unemployment benefits', 'Coronavirus recession', 'Stock market crash', 'Crisis' | Unemployment benefits', 'Unemployment', 'economic crisis', 'Crisis ', 'Recession' | 'Recession', 'Stock market crash', 'Financial crisis of 2007–2008', 'economic crisis', 'Unemployment benefits' |
| 2020Q3 | 'Unemployment benefits', 'Stock market crash', 'Abandoned pets', 'Mortgage calculator', 'Unemployment' | 'Government debt', 'Stock market crash', 'Mortgage loan', 'Unemployment', 'Coronavirus recession' | Unemployment benefits', 'Government debt', 'Gingival recession', 'Bankruptcy', 'Abandoned pets' |
| 2020Q4 | Gingival recession', 'Abandoned pets', 'Interest rate', 'Stock market crash', 'Interest' | 'Abandoned pets', 'Government debt', 'Stock market crash', 'Gingival recession', 'Agricultural Equipment' | 'Unwanted Body & Facial Hair Remova', 'Stock market crash', 'Abandoned pets', 'Mortgage law', 'Loanword' |
| 2021Q1 | 'Government debt', 'Pharmaceuticals & Biotech', 'Loanword', 'Investment fund', '1997 Asian financial crisis' | 'Vehicle Licensing & Registration', 'Mortgage law', 'Unemployment benefits', 'Fire & Security Services', 'Investment fund' | 'Pharmaceuticals & Biotech', 'Loanword', 'Vehicle Licensing & Registration', 'Investment', 'Mortgage law' |
| 2021Q2 | 'Pharmaceuticals & Biotech', 'Investment fund', 'Sports', 'Mortgage law', 'Baggage' | 'Pharmaceuticals & Biotech', 'Recruitment advertising', 'Sports', 'Maritime Transport', 'loan' | 'Pharmaceuticals & Biotech', 'Sports', 'Maritime Transport', 'Hand luggage', 'Construction Consulting & Contract' |
| 2021Q3 | 'Pharmaceuticals & Biotech', 'Return on investment', 'Commercial Lending', 'Sports', "Doctors' Offices" | 'Pharmaceuticals & Biotech', 'Hand luggage', 'Fuel Economy & Gas Prices', 'Return on investment', 'Hospitals & Treatment Centers' | 'Pharmaceuticals & Biotech', 'Gingival recession', 'Personal bankruptcy', 'Employment agency', 'Auto Financing' |
| 2021Q4 | 'Pharmaceuticals & Biotech', 'Maritime Transport', 'Aquaculture', 'Forestry', 'Industrial Materials & Equipment' | Pharmaceuticals & Biotech', 'Construction Consulting & Contract', 'Distribution & Logistics', 'Auto Financing', 'Maritime Transport' | 'Pharmaceuticals & Biotech', 'Fuel Economy & Gas Prices', 'Travel Agencies & Services', 'Entertainment Media', 'Vehicle Licensing & Registration' |

### The case of Vietnam

According to the report of General Statistics Offices[[5]](#footnote-5), in the first months of 2021, Vietnam's macro-economy continued to be stable and started to improve. Nevertheless, the wave of the Covid-19 pandemic breaking out in April has resulted in a dangerous situation. GDP growth rate in the 2021 Q3 was estimated at -6.17% over the same period last year. This decline is the deepest since Vietnam started to calculate and report publicly quarterly GDP. It led to 4.7 million people losing their jobs, 14.7 million people had to suspend production and business operations, 12.0 million people had their working hours cut and 18.9 million workers had their income reduced.

Figure 4.11 focuses on the fluctuation of the Weekly Tracker of Vietnam during the new wave of Covid-19. The pandemic consequences quite lead to the movement of the Tracker. The growth rate of new cases reduced the Weekly Tracker, while the growth rate of new deaths extremely hurt its value. The notable point is the trend of the Tracker in and out of the period Ho Chi Minh City was applied social-distancing and lock down directives. The Weekly Tracker pushes up to over 20% before the social-distancing period and immediately steep downed when the travel limitation is issued. A slightly downward trend appears during the locked down time. And when the locked down period stops, the trend of Weekly Tracker upgraded immediately.

Figure 4.12 provides insights on the contributions of variables to the model output in Vietnam. Economic Anxiety topics and categories had a large impact on the Weekly Tracker, but this impact comes from the downgrades of the svi's growth rate. It means that compared to the same period in 2020, the fear of people about the pandemic decreases. Different from last year, people kept calm during the Covid-19 new wave and, as the result, the economy was saved from the strong flop. In addition to the decrease in Economic Anxiety, the growth of Health Condition topics and categories had a large positive to the Weekly Tracker. It shows the adaptation and preparation of people for the pandemic, especially the peak of the Weekly Tracker during the locked down time was a significant impact of Health Condition. The adaptation and preparation are visible through the appearance of the Finance and Labor Market contributions, they showed the effort for increasing income during the shock time of people.

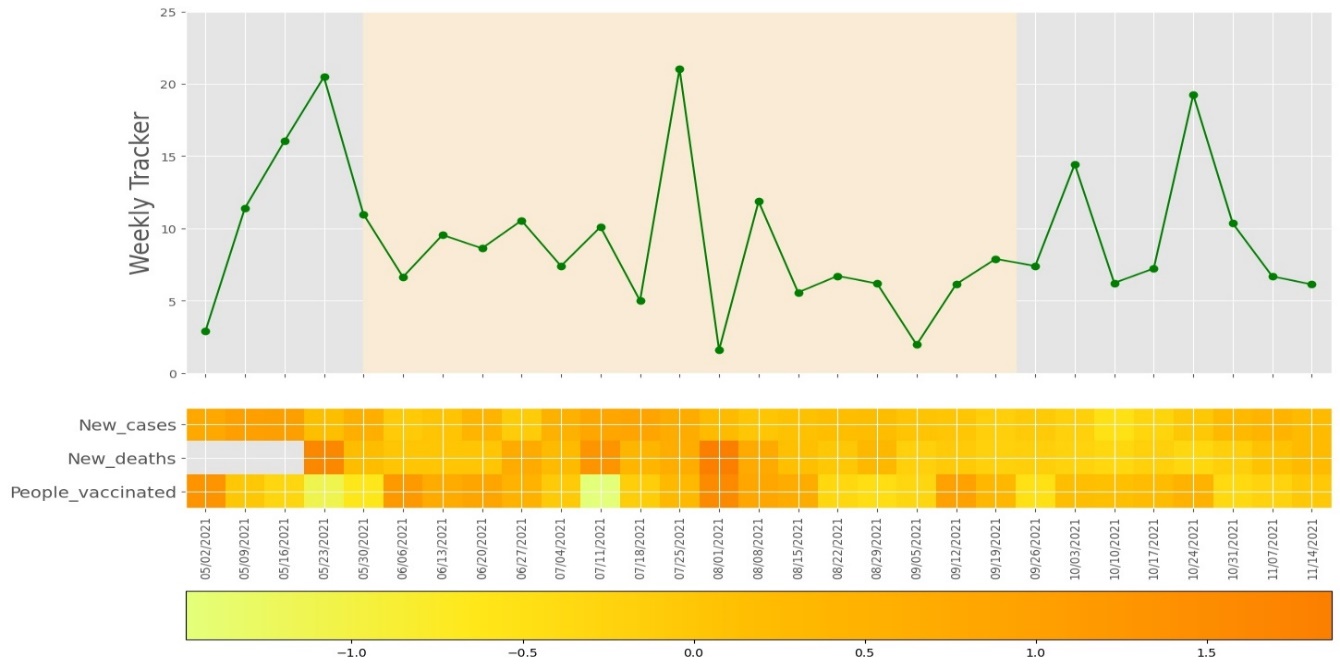


Figure .The Weekly Tracker of Vietnam during the new wave of Covid-19  
Note: This Figure shows the percentage change for the previous week of new cases and new deaths caused by the pandemic and the number of people vaccinated. The orange side in the Weekly Tracker graph covers months that Ho Chi Minh City applied to lock down protocol, from July to September 2021.  
Source: Calculated by the author   
Data: WHO

Figure 4.12 points out the important role of Consumption for the economy. The movement of the growth rate of topics and categories related to Consumption Services led to the trend of the Weekly Tracker. When the number of searches about Consumption is reduced, immediately the contribution of this variable decreases, and the Weekly Tracker downgrades. It leads to a downward trend during the social-distancing and locked down period. And when the travel limitation is stopped, the interest on Consumption increases, and the trend of the Weekly Tracker upgraded. It suggests that Consumption contributes largely to the recovery of the economy, so travel limitation protocols are carefully issued cause of avoiding negative impacts on the Consumption interest.



Figure . Drivers of the growth rate during the Covid-19 new wave in Vietnam   
Note: Contributions of variables to outcome is represented by Shapley Value. Search value index is indicated by the difference of svi year-over-year. Variable contributions and year-over-year differences svi are aggregated into economically relevant subgroups reflecting key economic sectors, after taking sum of all countries in sample.  
Source: Calculated by the author

# CONCLUSIONS AND POLICY IMPLICATIONS

## Conclusions

The result of this thesis shows the value of the Weekly Tracker for tracking GDP growth rate. As the findings of Baker & Fradkin (2017), Vosen & Schmidt (2011), Chen et al. (2014), Choi & Varian (2011), Balakrishnan & Kalpit Dixit (2013), and Martinez (2021), this thesis also considers the largely covering to economic sectors of Google Trends data relied on its big dataset. The output of the model support to **Stephens-Davidowitz & Varian** (2014), **Vosen & Schmidt (2011), Chen et al. (2014), Eraslan & Gotz (2020),** Lewis et al. (2020), Monache et al. (2021), Keane & Neal (2020) and Fetzer et al. (2020) about the usefulness of the Google Trends data for economic predicting activity. With high frequency, the Google Trends data has contributed to building a tracker index that can lead the quarterly GDP growth rate. The result points out, in order to construct a tracker for the economic activity, the ANN model outperforms the AR model for all periods, especially in the fluctuation time. During both the 2008 crisis and the Covid-19 pandemic in 2020, the median relative RMSE between the ANN model and the AR model is lower than 1, it means that the prediction of ANN model has lower error than the AR model, it is considered that the prediction of the model of this thesis is valuable. Especially, the relative value is lower than 1 across countries in the sample during the Covid-19 pandemic, it is the evidence proves that the ANN model outperforms the AR model in predicting economic activity during the coronavirus spread time. It is a major contribution of this thesis to find another approach for reacting well to the fluctuation period.

The Weekly Tracker suggests that the recovery of 22 economies in the sample has been constrained during the new wave of Covid-19. The growth rate estimate may be still positive, but the trend is horizontal, even in developed countries. Upper-middle-income countries are impacted negatively the most when a steep downtrend of Weekly Tracker has been immediately recorded since June. Almost all high-income countries in the sample find out the recovery in the first months of 2021, but it is cut soon in the third quarter of 2021, and the stagnation of growth rate may maintain to the end of 2021. The weekly GDP growth rate estimate was recorded close to 0% and even negative at some points. Compared to advanced countries and upper-middle-income countries, the Weekly Tracker of lower-middle-income economies was more stable during 2021 Q3 and Q4. The positive point is the impact of the Covid-19 during 2021 Q3 and Q4 seemly does not make a downward trend for the economic growth rate, as Weekly Tracker points out. The growth rate has remained at the value larger than 0%, not flop as the first wave of Covid-19. However, there is no evidence that the pandemic stops and it has no impact on the economies.

And with the Shapley value analysis, this thesis finds out some important sectors with have significant contributions to the growth rate during the pandemic. It is considered that Consumption topics and categories, especially trade and services, are the variables that have a key role in the recovery, it has a large contribution and positive impact on the model outcome. These results support the conclusion of Ankargren & Lindholm (2021), Woloszko (2020), Keane & Neal (2020). It is the same case for Labor Market that is believed that related to the worry of people during the first shock of Covid-19 (Eraslan & Gotz, 2020), but from the third quarter of 2020, it shows the change when the interest in topic "Jobs" and "Job interview" increases. By contrast, the finding is suitable to the conclusion of Fetzer et al. (2020) about the relation between this pandemic to the recession in the past. The topics and categories related to Economic Anxiety were popular. However, although Economic Anxiety remained popular searching and contribution to the Weekly Tracker, the search value index of topics and categories related to anxiety downgraded in 2021. The reason may be preparation with evidence is the growth of Health Condition queries. Finally, the results of this thesis also consider that the financial market is extremely important during the pandemic. People not only worry about their loans but also want to look up opportunities in the financial market.

## Policy implications

When studying the impact of independent variables on the model output, this thesis finds out some conclusions that are maybe helpful for policy implications.

Firstly, this thesis points out consumption is another sector that contributes largely to the economic recovery. During the pandemic, topics and categories related to consumption not only do not decrease but also increase extremely. Hence, this thesis found out evidences that the pandemic changes the behaviors of people significantly. During the Covid-19 crisis, people learn how to work, shopping, or pay bills from home. Life is dependable on the internet more than ever. It points out that, although the consumption demand still decreases because of the pandemic, the adaption of people leads to the change in consumption and consumption finds out the motivation to recover. It is the evidence that internet is the most important infrastructure that government has to focus on in the pandemic time. Hence, policies makers also concentrate services related to shopping online such as delivery service, banking service, education service, etc. Moreover, when production is stagnating, people find out another way to improve their wealth by participating in the financial market. Policies related to the financial market is may be considered at this time. Another importance is cybersecurity.

Secondly, besides consumption, the results show the large contribution of labor market and trading for the recovery after the depression because of the Covid-19 pandemic. Not only they are the main sectors of the economy but also they are economic representatives. The high growth of labor and trading lead to high economic growth, low growth of labor and trading lead to the flop of the economic growth. At the beginning of the pandemic, production and trading faces a considerable flop, the motivation of the economic growth is taken off, so the fall of the growth rate appears in the second quarter of 2020. However, at the end of 2020 and in 2021, production and trade are allowed back to activate, they get an impressive growth and lead the economy to recover. This point catches on the trade-off between limiting the spread out of coronavirus and the economy. When the Covid-19 pandemic outbreaks, the major economies decide to use the policies related to limitation travel, social distancing, and even lockdown to reduce the risk of the virus spreading out. However, these policies constrain production and trade activities and lead to downward economic growth. Until these policies are scaled back at the end of 2020, the economies find out the recovery based on the comeback of production and trade. Therefore, policies makers need to be careful when applying the travel limitation protocol during the pandemic.

Eventually, the fear of people during the pandemic has an extreme impact on the growth rate. During the pandemic, people relate the situation to the previous recession, they worry about their jobs, loan, or financial assets. And even in the time of economic recovery, the fear about the crisis remains largely. They worry that the pandemic may come back, and they try to find solutions to prepare well if facing an extreme depression. Comforting residents is an activity that government should consider and improve their knowledge about the pandemic is also important.

## Limitations

This thesis also has some limitations. Combining monthly official variables and weekly Google Trends indicators brings the opportunities to compare the contribution of them for the prediction, but it leads to the limit about the real-time implications, and mistakes from converting monthly data to weekly data. The data from 22 economies from Asia and Pacific is still small and needs to be expanded to more countries. On the other hand, the heterogeneity is large despite the small number of countries. There is a large number of topics and categories not strong link to the economic sectors, and monthly official variables need more literature review to improve their reliability. Eventually, it is needed to have other tests to evaluate the valuation of the prediction of the mix-frequencies model.

## Recommendations for future work on this topic

Further research can include official weekly data to study the impact of different sectors (such as financial variables, commodity prices, or electricity consumption) on the GDP growth rate when the second wave of coronavirus widespread has hurt globally. Another approach is combining more monthly variables to compare the reaction of official indicators and Google search to the fluctuation of the economy during the Covid-19 pandemic.

# REFERENCES:

Ankargren & Lindholm (2021). *Nowcasting Swedish GDP Growth Data* (Working Paper No 154). Konjunktur Institutet

Baker, S., & Fradkin, A. (2017). The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. *The Review Of Economics And Statistics*, 99(5), 756-768. https://doi.org/10.1162/rest\_a\_00674

Baldwin, R. & Mauro, W. (2020). *Economics in the Time of Covid-19*. CEPR Press.

Csáji, B. (2001), *Approximation with Artificial Neural Networks*, Faculty of Sciences, Etvs Lornd University, Hungary

Combes, S., & Bortoli, C. (2016). Nowcasting with Google trends, the more is not always the better*. Proceedings of the 1st International Conference on Advanced Research Methods and Analytics*. https://doi.org/10.4995/carma2016.2016.4226

Chen, T., So, E. P., Wu, L., & Yan, I. K. (2014). THE 2007-2008 U.S. Recession: What did the Real-time Google Trends data tell the united states? *Contemporary Economic Policy, 33*(2), 395–403. https://doi.org/10.1111/coep.12074

CHOI, H. Y. U. N. Y. O. U. N. G., & VARIAN, H. A. L. (2012). Predicting the present with google trends. *Economic Record*, 88, 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Eraslan, S., & Götz, T. (2021). An unconventional weekly economic activity index for Germany. *Economics Letters*, 204, 109881. https://doi.org/10.1016/j.econlet.2021.109881

Fetzer, T., Hensel, L., Hermle, J., & Roth, C. (2020). Coronavirus perceptions and economic anxiety. *The Review of Economics and Statistics*, 1–36. https://doi.org/10.1162/rest\_a\_00946

Ficetola, G. (2013). Is interest toward the environment really declining? The complexity of analysing trends using internet search data. *Biodiversity And Conservation*, *22*(12), 2983-2988. https://doi.org/10.1007/s10531-013-0552-y

Gómez, M., Orden-Cruz, C., & Prado, R. (2021). Google trends as predictor of grain prices. *Ekonomika Poljoprivrede, 68*(1), 203–211. https://doi.org/10.5937/ekopolj2101203g

Giannone, D., Lenza, M., & Primiceri, G. (2018). Economic Predictions with Big Data: The Illusion of Sparsity. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3166281

Huyen, N. (2021). Report on the Covid-19 impacts on labour and employment situation in the third quarter of 2021. Retrieved 17 December 2021, from https://www.gso.gov.vn/en/data-and-statistics/2021/10/report-on-the-covid-19-impacts-on-labour-and-employment-situation-in-the-third-quarter-of-2021/

Jahn, M. (2020). Artificial neural network regression models in a panel setting: Predicting economic growth. *Economic Modelling*, 91, 148–154. https://doi.org/10.1016/j.econmod.2020.06.008

Jena, P. R., Majhi, R., Kalli, R., Managi, S., & Majhi, B. (2021). Impact of covid-19 on gdp of major economies: Application of the artificial neural network forecaster. *Economic Analysis and Policy*, 69, 324–339. https://doi.org/10.1016/j.eap.2020.12.013

Keane, M. P., & Neal, T. (2020). Consumer panic in the covid-19 pandemic. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3600018

Lewis, D., Merterns, K., Stock, J. & Trivedi, M. (2020). *Measuring real activity using a weekly economic index (Staff Reports No 920)*. Federal Reserve Bank of New York.

Loermann, Julius and Maas, Benedikt (2019): *Nowcasting US GDP with artificial neural networks.*

Lundberg, S., & Lee, S. (2021). A Unified Approach to Interpreting Model Predictions. Retrieved 17 December 2021, from https://arxiv.org/abs/1705.07874

Mccallum, M., & Bury, G. (2013). Google search patterns suggest declining interest in the environment. *Biodiversity And Conservation, 22*(6-7), 1355-1367. https://doi.org/10.1007/s10531-013-0476-6

Molnar, C. (2020), Interpretable Machine Learning, https://books.google.fr/books?hl=en&lr=&id=jBm3DwAAQBAJ&oi=fnd&pg=PP1&dq=Interpretable+Machine+Learning.+A+Guide+for+Making+Black+Box+Models+Explainable.&ots=EfAX%20iFDY3&sig=jPrA7RHxQ5j7Bura6KZnPCMZrj8&redir\_esc=y#v=onepage&q=Interpretable%20Machine%20Learning.%20A%20Guide%20for%20Making%20Black%20Box%20Models%20%20Explainable.&f=false (accessed on 28 August 2020).

Monache, D., Emiliozzi, S. & Nobili, A. (2021). *Tracking economic growth during the Covid-19: A weekly indicator for Italy*. Banca D'Italia.

Muchisha, N. D., Tamara, N., Andriansyah, A., & Soleh, A. M. (2021). Nowcasting Indonesia’s gdp growth using machine learning algorithms*. Indonesian Journal of Statistics and Its Applications*, 5(2), 355–368. https://doi.org/10.29244/ijsa.v5i2p355-368

SANUSI, N. A., MOOSIN, A. F., & KUSAIRI, S. (2020). Neural network analysis in forecasting the MALAYSIAN GDP. *The Journal of Asian Finance, Economics and Business*, 7(12), 109–114. https://doi.org/10.13106/jafeb.2020.vol7.no12.109

Suhoy, T. (2009). Query indices and a 2008 downturn. *Discussion Paper No. 6*, Bank of Israel, Israel. Swamy, A., & Yerpude, A. (2016).

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal Of Forecasting*, 30(6), 565-578. https://doi.org/10.1002/for.1213

Willard, S., & Nguyen, M. (2013). Internet Search Trends Analysis Tools Can Provide Real-time Data on Kidney Stone Disease in the United States. *Urology, 81*(1), 37-42. https://doi.org/10.1016/j.urology.2011.04.024

Woloszko, N. (2020). *Tracking activity in real time with Google Trends* (Working Paper No 1634). OECD.

# APPENDIX

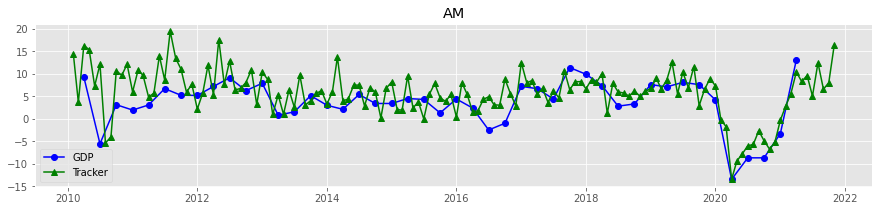
**APPENDIX A**

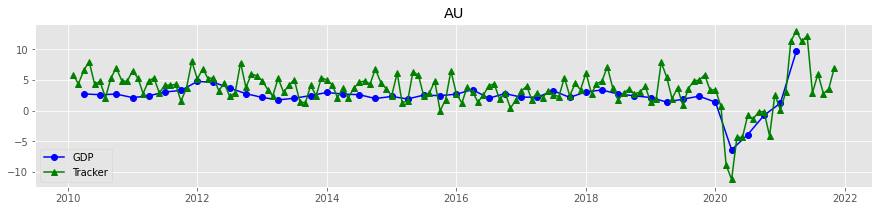
**Table of data description:**

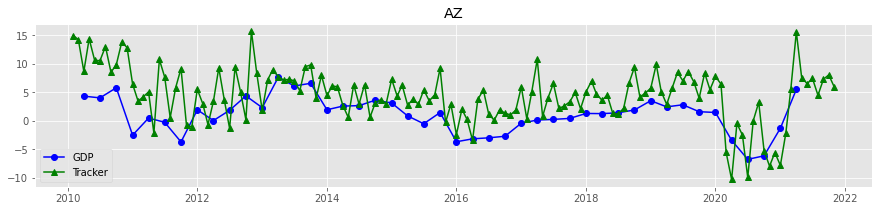
|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Indicators | Source | Papers |
| Output training variable (quarterly frequency) | | | |
| GDP growth rate | GDP growth rate year-over-year | ADB | Monache et al.(2021); Jardet & Meunier( 2020); Lewis et al.(2020); Woloszko (2020); Tamara et al. (2020); Zhemkov (2021) |
| Google Trends estimators (search topics and categories, monthly frequency for training and weekly frequency for predicting) | | | |
| Economic Anxiety | 'United States housing bubble', 'Deed in lieu of foreclosure', 'Bureau of the Fiscal Service', 'Fuel Economy & Gas Prices', 'Government', "2010 United States foreclosure crisis", "economic crisis", "Crisis 1", "1997 Asian financial crisis", 'National debt of the United States', 'Government debt', "Crisis hotline", "Crisis 2", "Recession", 'Gingival recession', 'Force', "Financial crisis", "Gingival recession", "Coronavirus recession", "Global recession", "Stock market crash", 'Insurance','Mergers, Acquisitions, and Other Restructuring Activities: An Integrated Approach to Process, Tools, Cases, and Solutions', 'Foreclosure', "Unemployment", "Unemployment benefits","Unemployment", "Natural rate of unemployment", "Unemployment benefits in Texas", "Unemployment benefits in France", "A World Without Work: Technology, Automation, and How We Should Respond", "Unemployment insurance in Japan", 'Bankruptcy', 'Personal bankruptcy', 'Bankruptcy proceedings', 'Bankruptcy' | Google Trends | Woloszko (2020); Choi & Varian (2011); Suhoy (2009); Martinez et al. (2021) |
| Health Condition | 'Strength training', 'Pharmaceuticals & Biotech', 'Unwanted Body & Facial Hair Remova', 'Medical Facilities & Services', 'Mental Health', 'Hospitals & Treatment Centers', "Doctors' Offices", 'Health Insurance', 'Pharmacy' | Google Trends |
| Consumption Services | 'Entertainment Media','Abandoned pets', 'Chapter 7, Title 11, United States Code', 'Online banking', 'Alcoholic Beverages', 'Boats & Watercraft' , 'Fire & Security Services','Emergency Services', 'Restaurants', 'Entertainment Industry', 'Performing Arts', 'Birthday', 'Social Services', 'Travel', 'Sports', 'Autos & Vehicles', 'Home & Garden', 'Events & Listings', 'Vehicle Licensing & Registration', 'Home Appliances', 'Home Furnishings', 'Footwear', 'Hotels & Accommodations', 'Travel Agencies & Services', 'Apparel', 'Legal Services', 'Education' | Google Trends |
| Consumption Goods | 'Consumer Electronics', 'Gifts & Special Event Items', 'Hand luggage', 'Bag tag', 'Baggage', 'Luxury Goods', 'Food & Drink', 'Vehicle Brands', 'Birthday card', 'Shopping', 'Tobacco Products', 'Grocery & Food Retailers', 'Veterinarians', 'Computers & Electronics''Pharmacy', 'Carpooling & Ridesharing', 'Sports', 'Animal Products & Services', 'Fitness', 'Weddings', 'Car Rental & Taxi Services', 'Autos & Vehicles', 'Tourist Destinations', 'Home & Garden', 'Events & Listings',  'Grocery & Food Retailers', 'Vehicle Licensing & Registration', 'Timeshares & Vacation Properties', 'Home Appliances', 'Mass Merchants & Department Stores', 'Car Electronics', 'Fashion & Style', 'Trucks & SUVs',  'Home Furnishings', 'Footwear', 'Cruises & Charters', 'Hotels & Accommodations', 'Luggage & Travel Accessories', 'Fast Food', 'Book Retailers', 'Veterinarians', 'Spas & Beauty Services', 'Acting & Theater',  'Travel Agencies & Services', 'Apparel', 'Legal Services', 'Education','Computers & Electronics' | Google Trends |
| Labour Market | 'Civil service entrance examination', 'Job interview', 'Job', 'Temporary Solutions', 'Employment agency', 'Recruitment', 'Military recruitment', 'Recruitment advertising', 'Recruitment & Staffing', 'Developer Jobs', 'Job Search', 'Jobs' | Google Trends |
| Housing Construction | 'Building Materials & Supplies', 'Real Estate Agencies', 'Real Estate', 'Affordable Housing 2','Affordable Housing 1', 'House price index', 'Home Improvement', 'Construction Consulting & Contract', 'Civil Engineering', 'Construction & Maintenance' | Google Trends |
| Finance | 'The Poor Had No Lawyers: Who Owns Scotland and How They Got it', 'Professional & Trade Associations', 'Investment', "investment", 'Return on investment', 'Investment fund' , 'Finance', "Mortgage loan",'Student loan', 'Credit & Lending', 'Loan', 'Interest', 'Interest rate', 'Mortgage law', 'Mortgage calculator', 'Fixed-rate mortgage', 'Auto Financing', 'College Financing', 'Home Financing', "loan", "Loanword", "Student Loan Fund" | Google Trends |
| Business Services | 'Retail Trade','Rail Transport', 'Office Space', 'Maritime Transport', 'Computer Hardware', 'Advertising & Marketing', 'Architecture', 'Business Services', 'Printing & Publishing', 'Distribution & Logistics', 'Export', 'Import & Export', 'commercial building', 'Commercial Lending', 'Internet & Telecom', 'Office Supplies', 'Programming' | Google Trends |
| Industrial | 'Agricultural Equipment', 'Industrial Materials & Equipment', 'Aquaculture', 'Forestry', 'Agriculture & Forestry', 'Aviation', 'Chemicals Industry', 'Food Production', 'Freight & Trucking', 'Transportation & Logistics', 'Mail & Package Delivery' | Google Trends |

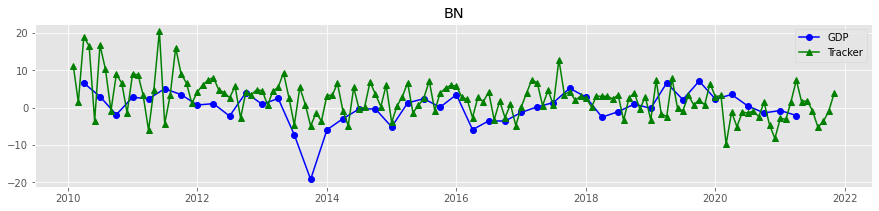
**APPENDIX B**

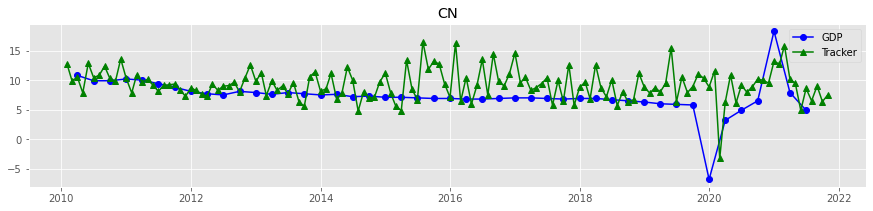
**The prediction of ANN model**

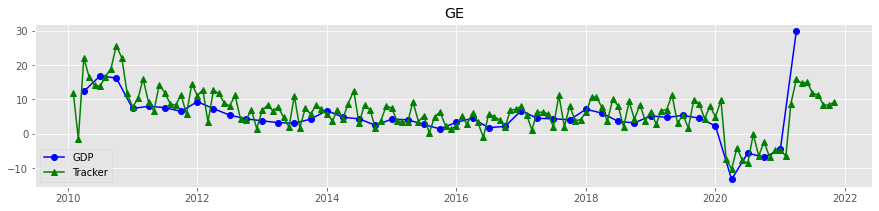


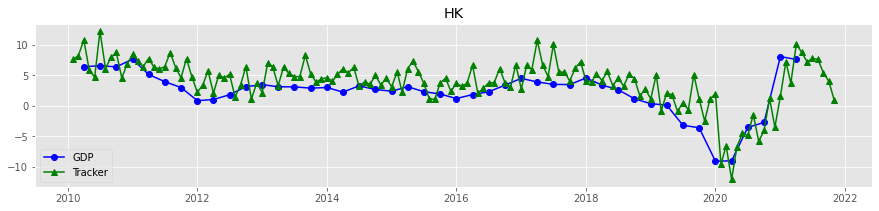


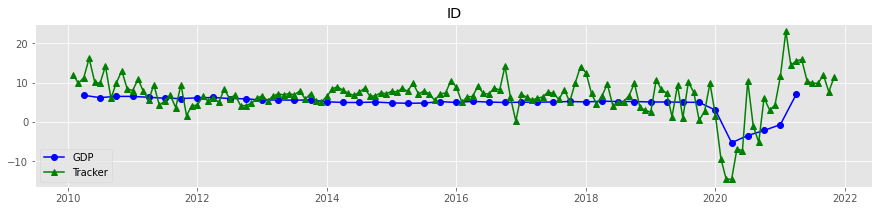


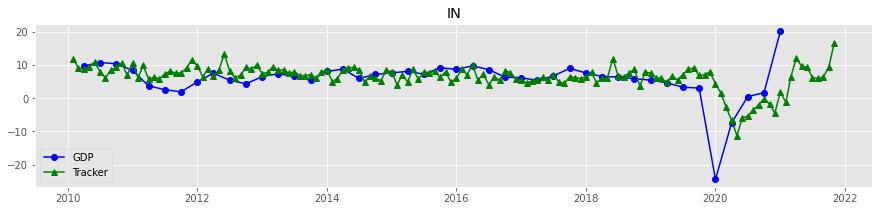


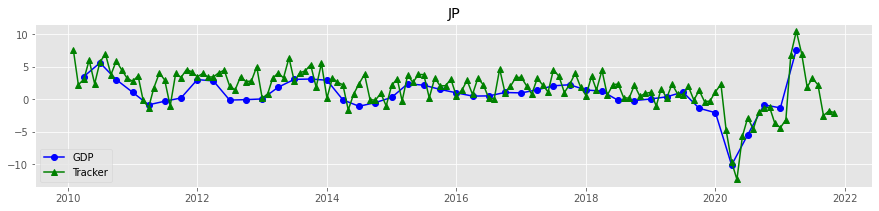


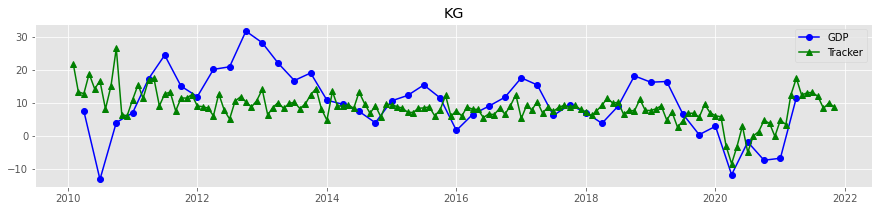


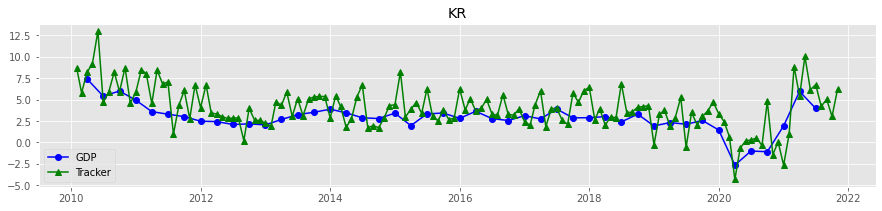


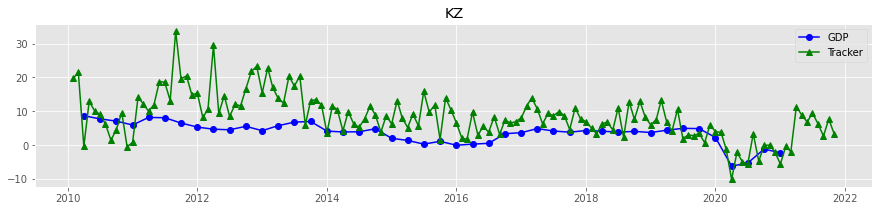


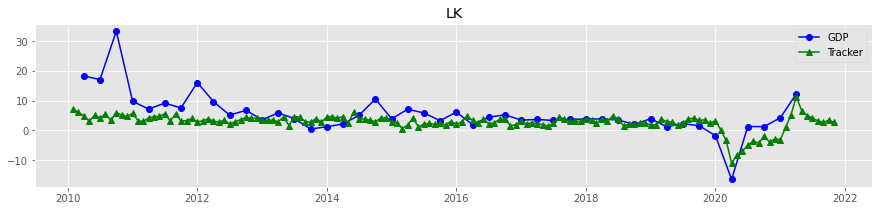


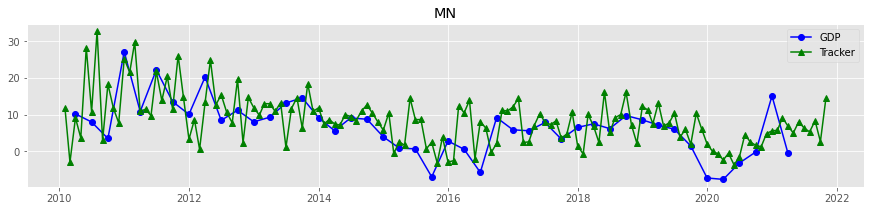


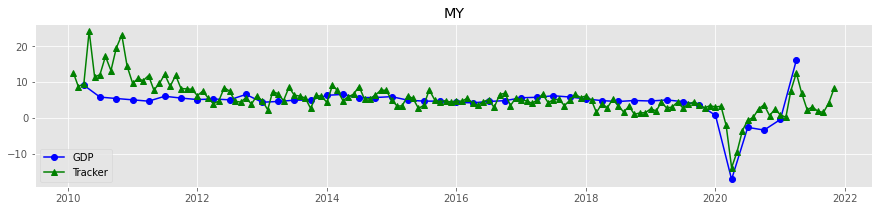


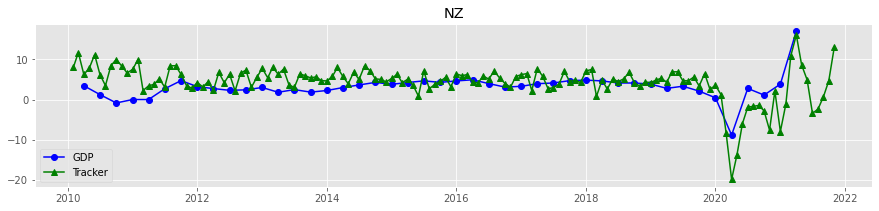


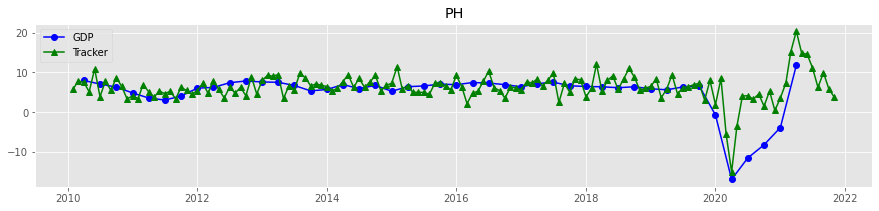


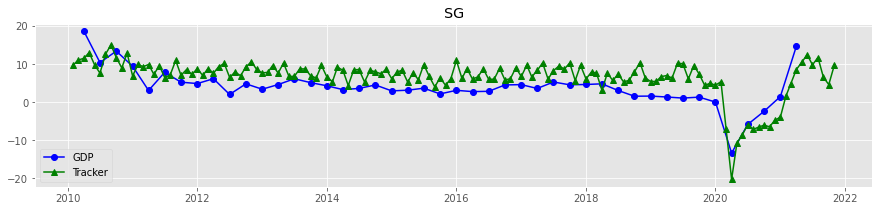


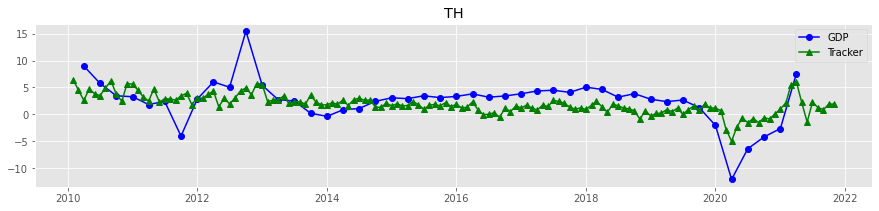


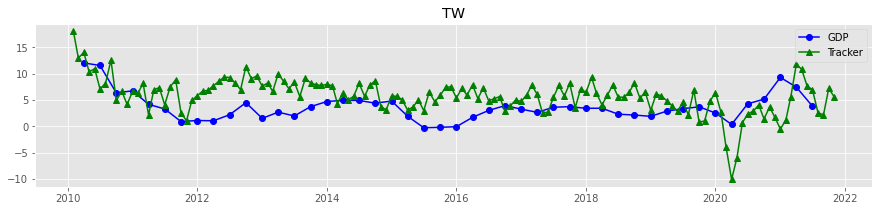


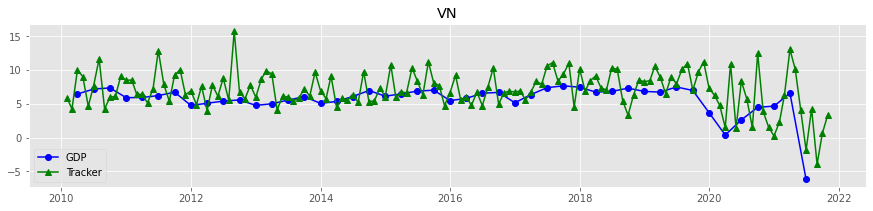






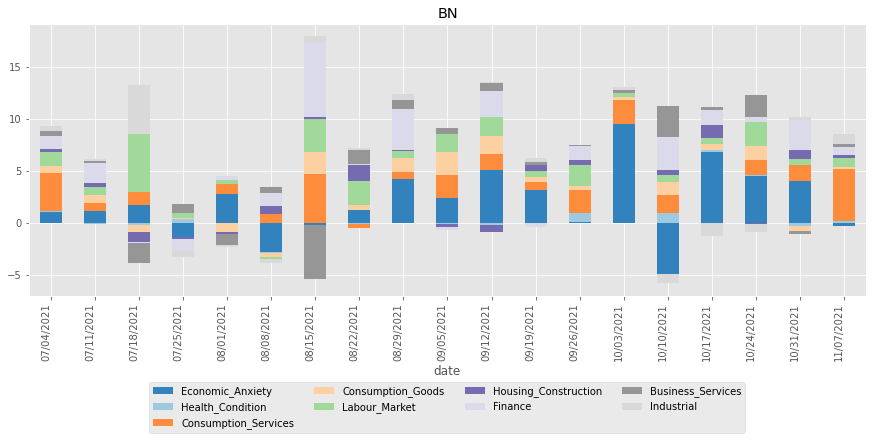




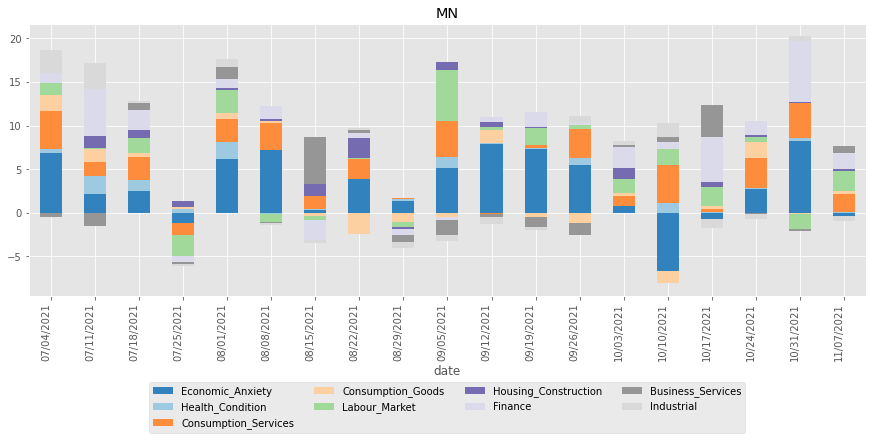


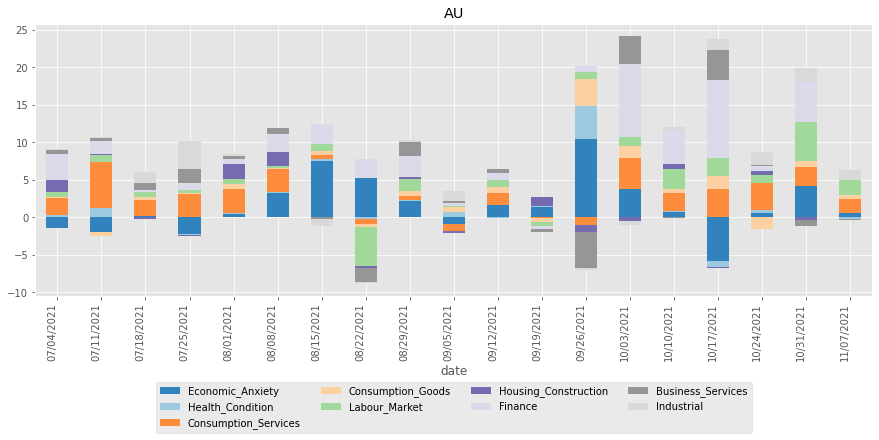
**APPENDIX C**

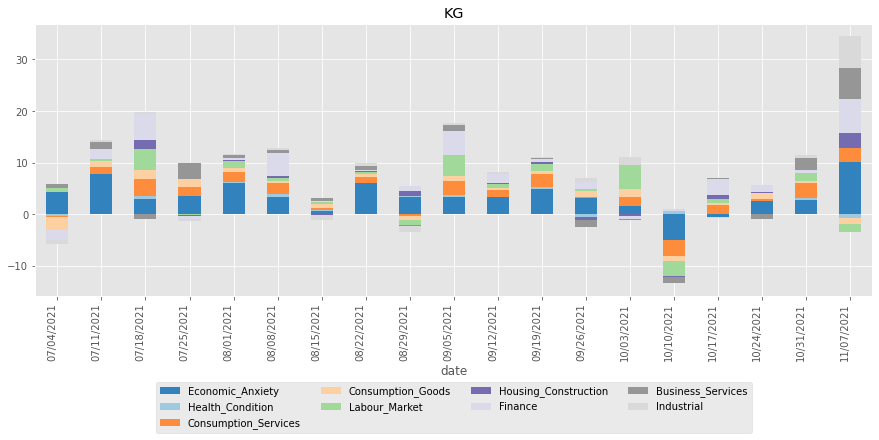
**Weekly Shapley value**

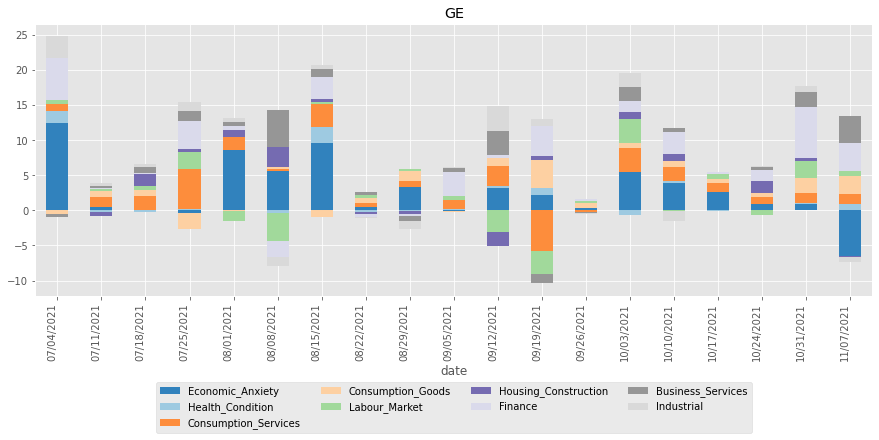






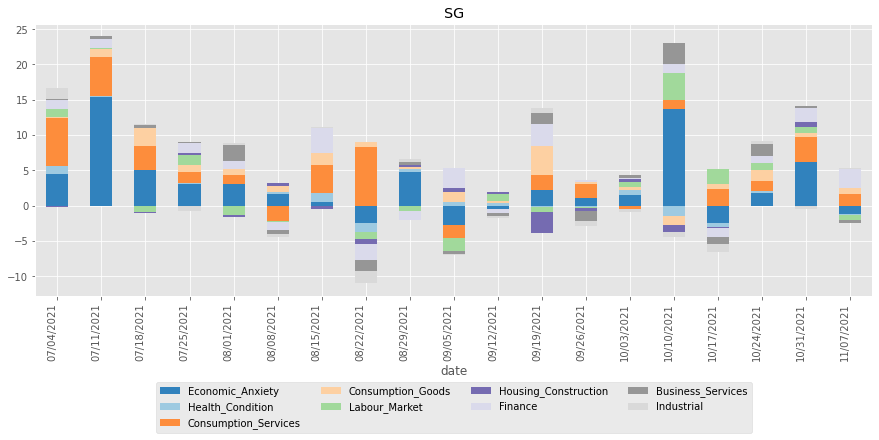


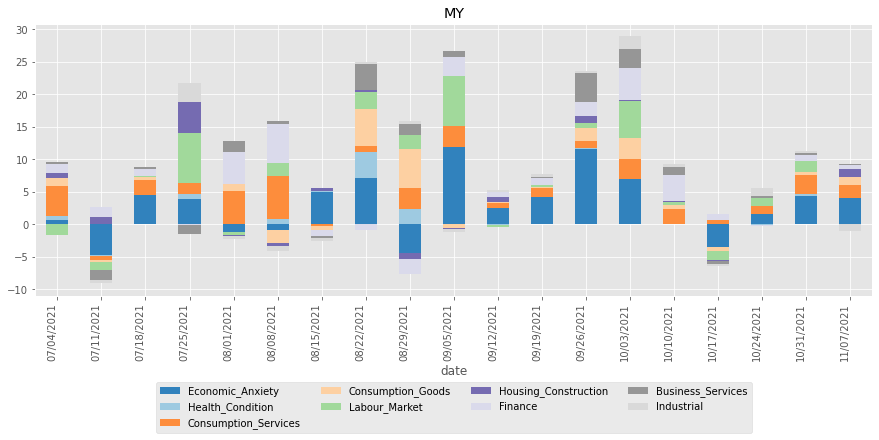






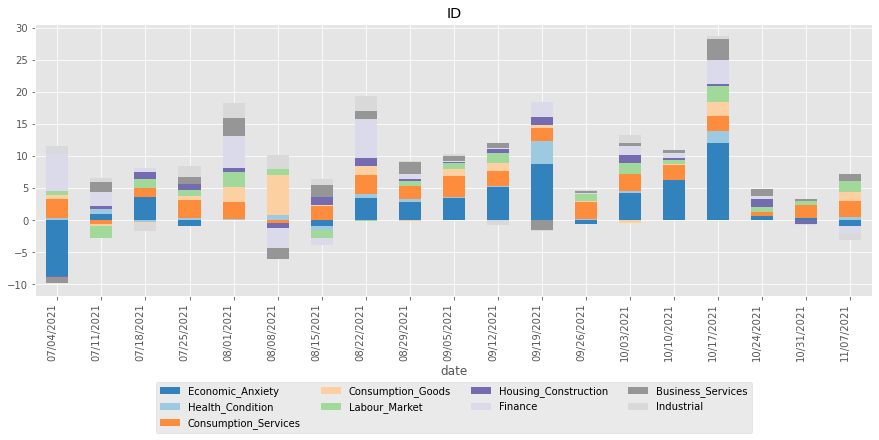




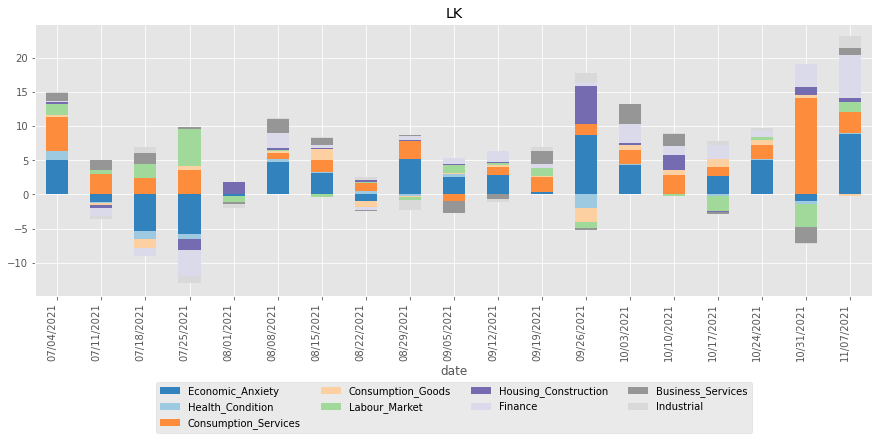


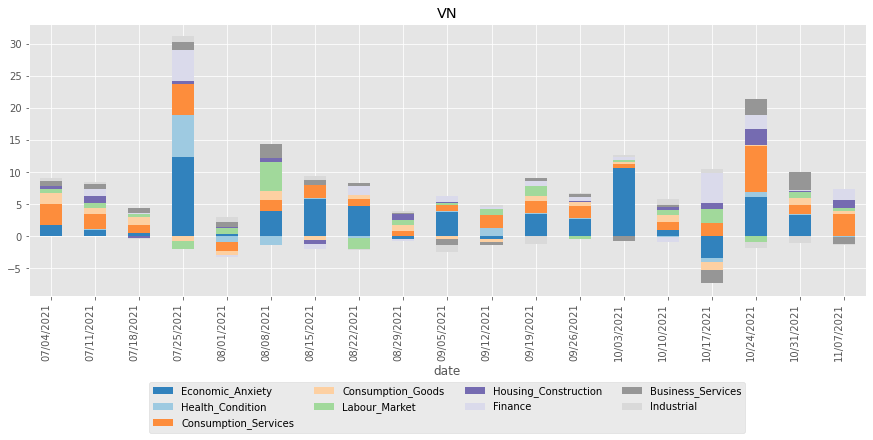


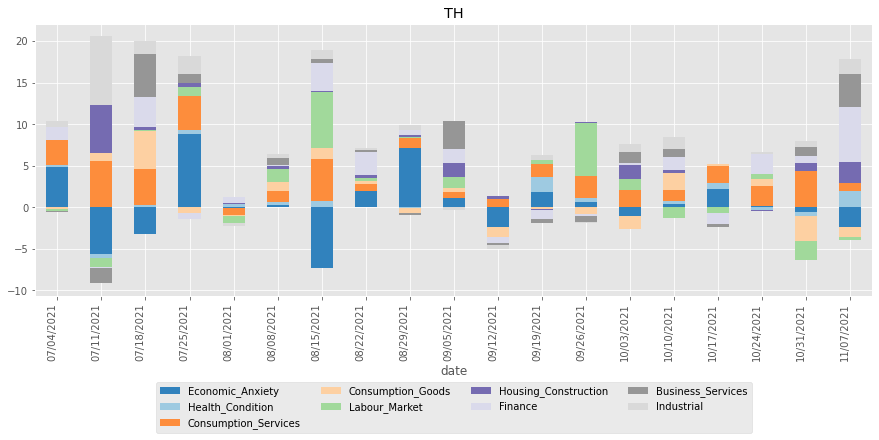








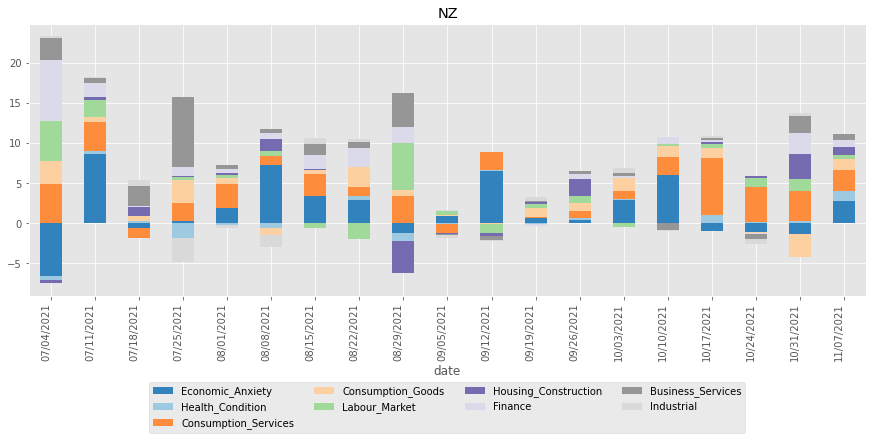


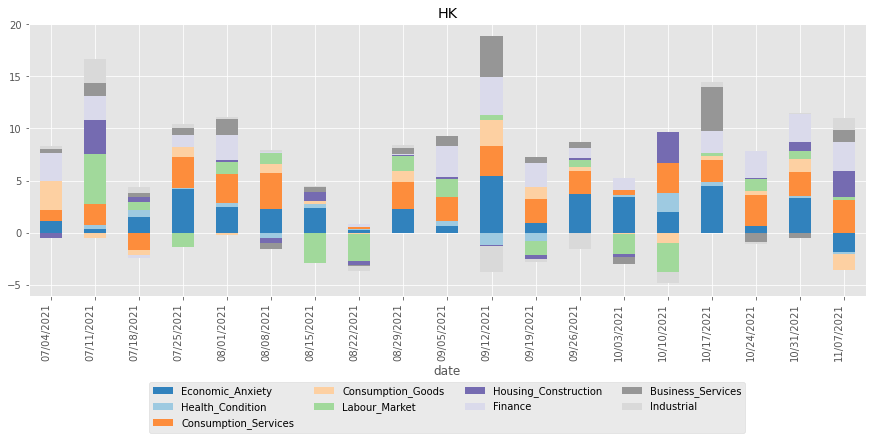












1. Asian Development Outlook 2021 Update Highlights [↑](#footnote-ref-1)
2. https://support.google.com/trends/answer/4365533?hl=en [↑](#footnote-ref-2)
3. https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories [↑](#footnote-ref-3)
4. "Black box" describes the situation that machine learning algorithm is so complex that the produced outcome can be not explained, despite it proves accurate. [↑](#footnote-ref-4)
5. Report on the covid-19 impacts on labour and employment situation in the third quarter of 2021 (Source: General Statistics Offices) [↑](#footnote-ref-5)