PIER

PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Predicting the Present Revisited: The Case of Thailand

by

Voraprapa Nakavachara and Nuarpear Warn Lekfuangfu

October 2017

Discussion Paper

No. 70

Predicting the Present Revisited: The Case of Thailand

Voraprapa Nakavachara[a,*] and Nuarpear Warn Lekfuangfu[a,b]

[a] Faculty of Economics, Chulalongkorn University, Phayathai Road, Pathumwan, Bangkok 10330, Thailand
[b] Centre for Economic Performance, London school of Economics and Political Science, London WC2A 2AE, United Kingdom

Version: 20 September 2017

Abstract

Google is currently the most-used search engine in the world. There are approximately 3.5 billion searches being conducted on Google each day. With real-time processing, Google Trends data can be used in a prediction technique called nowcasting (or "predicting the present") – using the current period's real-time information to estimate the current period's indicators of interest. In this paper, we showed how Google Trends can be used for nowcasting Thailand's various economic indicators. The sectors being analyzed are (i) the labor market sector (unemployment rate and unemployment registration), (ii) the real sector (automobile sales), and (iii) the financial sector (SET index). The results revealed that incorporating the Google Trends data into the prediction models improved the Adjusted R-Squared and improved the predication accuracies under various measures.

Keywords: Nowcasting, Google Trends

JEL Codes: J01, L62, G10, G17

---

[*] Corresponding author.
Email addresses: voraprapa.n@chula.ac.th (V. Nakavachara), np_warn@yahoo.com (N.W. Lekfuangfu)
The authors would like to thank Supisara Reungpeerakul for her valuable research assistance. The authors also would like to thank Piti Disyatat, Nasha Ananchotikul, Piriya Pholphirul, Attakrit Leckcivilize, and Paphatsorn Sawaengsuksant for their helpful comments on an earlier version of this paper. The results and findings reflect the personal view of the authors and do not reflect the views of their professional institutions.

1. Introduction

As of 2017, Google is the most-used search engine in the world, accounting for at least 79% of the world's internet search traffic.[1] There are approximately 3.5 billion searches being conducted on Google per day.[2] Globally, there are currently 3.8 billion internet users. The world's internet penetration rate, defined as percentage of the population using internet divided by the total population, is approximately 50%. The number of internet users have grown 10% from the previous year. Regarding the mobile devices, there are currently 4.9 billion unique mobile device users worldwide.[3] With rapidly increasing internet population and thus Google users, it is worthwhile examine the information of the searches that Google collects and determine how it can help provide insights into various topics that are of public interests.

Scholars have tried to study and utilize the search information that Google collected in research. In particular, Choi and Varian (2009a, 2009b, 2012) and McLaren and Shanbhogue (2011), used Google Trends – Google's search volume index indicating how often a term or a phrase has been searched by internet users relative to other terms or phrases over a period of time – in predicting various economic indicators such as automobile sales, home sales, travel volume, consumer confidence, unemployment rate, and initial claims for unemployment insurance.

One of the most important advantages of Google Trends is that it is updated almost on a real-time basis. Once a new search is conducted, such search information is collected and then later used to compute the Google Trends data. Thus, the real-time aspect of Google Trends

[1] Search Engine Market Share. Retrieved from https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0&qptimeframe=Y (as of 26 July 2017).
[2] Google Search Statistics. Retrieved from http://www.internetlivestats.com/google-search-statistics/ (as of 28 July 2017).
[3] We Are Social (2017). Digital in 2017: Global Overview. Retrieved from https://wearesocial.com/special-reports/digital-in-2017-global-overview (as of 16 August 2017)

is useful in a prediction technique called nowcasting. One can think of nowcasting as an improved version of forecasting. Choi and Varian (2009a, 2009b, 2012) explained nowcasting as “predicting the present.” While forecasting means using the previous period’s data to predict the following period’s economic indicators, nowcasting simply means using the current period’s data to predict the current period’s economic indicators.

The literature documented successful attempts in using Google Trends to improve the predictions of many economic indicators in various countries. Since the current-period’s Google Trends information can be retrieved almost real-time, much earlier than the time that current period’s economic indicators actually come out, incorporating Google Trends information into the nowcasting model can improve the predictions. Choi and Varian (2009a, 2009b, 2012), Askitas and Zimmermann (2009), Suhoy (2009), McLaren and Shanbhogue (2011), Carriere-Swallow & Labbe (2013), Fonduer & Karame (2013), Vincente, Lopez-Menendez, and Perez (2015), and Seabold and Coppola (2015) are among the literature that illustrated such prediction methodologies and how Google Trends improved the outcomes.

However, nowcasting using Google Trends does have some drawbacks. Google search volume indices, although reveal public interests at the time of the search, do not always reflect the actions that people will actually take. The fact that people conduct a Google search can only be interpreted as their desire to acquire more information on the subject. It does not reveal their opinions about the subject. Thus, the correlations of Google Trends with the actual economic indicators could be noisy. In addition, Google does not reveal the exact methodology that it uses in calculating the Google search volume index. Thus, researchers can never cross-check the calculation and will have to rely on Google for the accuracy and the consistency of the data. Despite the drawbacks, Google Trends still provides useful real-time information and thus improves the predictions of many economic indicators.

As already discussed, the number of internet users have grown 10% globally. The growth rate is highest, at 15%, for the Asia-Pacific region.[4] Within the Asia-Pacific region, Thailand is one of the countries with high internet penetration rate of 67%.[5] The country currently has 46 million internet users, 21% increase from the previous year. There are currently 47.9 million unique mobile device users. Approximately 11.58 million people reported having conducted an online purchase and the country's total revenue of e-commerce market (in 2016) was USD 2.8 billion.[6] With significant volume of internet activities, Thailand would make an interesting case study. For emerging middle-income countries, only a few studies have explored the potential of Google Trends in predicting the economic indicators. Carriere-Swallow and Labbe (2013) studied the role of Google Trends in nowcasting the automobile market in Chile. For Turkey, Chadwick and Sengul (2012) studied how Google Trends can help predict the country's unemployment rate and Zeybek and Ugurlu (2015) studied how Google Trends can help predict the country's credit demand. Seabold and Coppola (2015) explored how Google Trends can help predict the price levels in Costa Rica, El Salvador, and Honduras. For Thailand, to the best of our knowledge, besides an earlier version of this paper (Lekfuangfu, Nakavachara, and Sawaengsuksant (2016)), there is currently no other paper on how Google Trends can improve the predictions of Thailand's economic indicators. Therefore, this paper intends to bridge this gap.

In this study, we show how Google Trends can be used to nowcast Thailand's various economic indicators. We focus our analyses under three sectors, namely, (i) the labor market sector, (ii) the real sector, and (iii) the financial sector. The paper is organized as follow. Section 2 discusses Google Trends and how Google Trends emerged over time. Section 3 uses the

---

[4] Tie with the Middle East.
[5] Other countries with high internet penetration rates are Brunei (86%), Singapore (82%), and Malaysia (71%).
[6] We Are Social (2017). Digital in 2017: Southeast Asia. Retrieved from https://wearesocial.com/special-reports/digital-southeast-asia-2017 (as of 16 August 2017)

Google Trends data to nowcast Thailand's economic indicators. The econometrics models and the results are discussed under this section. Section 4 concludes the paper and discusses the authors' viewpoint regarding the future of economics research under the open data environment.

2. Evolution of Google Trends

Google Trends, first launched in 2004, is a web service by Google that reports trends of search keywords being conducted on Google. Specifically, Google Trends reports a search volume index – indicating how often the keyword was searched relative to the total number of searches from the same time/location. The index is normalized to be 0 to 100 over the selected time period. One can retrieve the search volume index data dated back to January 2004.[7]

Initially, the Google Trends data can be retrieved on a weekly frequency dated back to January 2004. (Other frequency types can be retrieved but with a shorter time span.) In 2009, Hal Varian, the chief economist of Google, first wrote papers on how Google Trends can be used to nowcast economic indicators (see Choi and Varian (2009a, 2009b, 2012)) by utilizing the weekly frequency version of Google Trends to nowcast monthly economic indicators. The fundamental of nowcasting using Google Trends is that Google Trends data come out more frequently and sooner than the actual official economic indicators. Therefore, the current period's Google Trends data are usually already available and can be used in the prediction of the current-period economic indicators (which usually come out later after the period has

[7] The history of the internet could be traced back to around 1960s when computers were connected for the first time and the first message was being sent between them. However, it was not until early 1990s when the internet was made available to general public. Back then it was difficult for people to find the information they wanted from the internet. Therefore, in 1994, Jerry Yang and David Filo created a web directory search that eventually became Yahoo. Many other search engines were created after that, including Google which was founded in 1998 by Larry Page and Sergey Brin. Google became more popular and outperformed other search engines due to its clean and simple user interface and efficient search algorithm. As of 2017, Google is currently the most-used search engine in the world with approximately 3.5 billion searches being conducted each day.

ended). Since then, many studies (see the previous section) have followed their methodology in using the weekly Google Trends data to nowcast the monthly economic indicators.

As of mid 2016, Google changed how Google Trends data are released to the public. The default frequency type that Google releases to the public when ones retrieve the data dated back to January 2004 is monthly instead of weekly. (Other frequency types can be retrieved but with a shorter time span.) However, the current-period monthly Google Trends data are readily available at the beginning of the month and the data is updated almost on a real-time basis, as the searches are being conducted, throughout the month. Thus, the fundamental of nowcasting using Google Trends is still valid since the current-period Google Trends data can be retrieved at any point in time during the period and much earlier than the time the actual economic indicators become available.

With 67% internet penetration rate, Thailand would make an interesting case study on how search data like Google Trends can reflect the public interests that perhaps could translate into people behaviors. To quickly illustrate the point, Figure 1 shows the Google Trends data for the keyword "หวย" (an informal Thai word for bi-monthly state lottery draw) over the 90-day period in which the daily data can be retrieved, restricting the location of the search to be from Thailand. Figure 1 provides supportive evidence for the coexisting of people's online search behavior and their real-world activities on two accounts. First, the co-movement of both trendlines shows that people's online search behavior corresponds their real-world actions in real time. To be specific, both trendlines peak on the 1st and the 16th of each month. The dates correspond to the dates that Thailand's Government Lottery announces its winners. Second, the accessibility of the search engines, especially Google, is not restricted to just the more sophisticated group of people in the Thai society. Purchasing of state lottery, exceedingly popular with approximately 71 million tickets being issued each round, is highly concentrated among lower socio-economic class. Therefore, the behavior captured in Figure 1 suggests that online search engine is widely used across social spectrum in Thailand.

3. Nowcasting Thailand

This section will demonstrate how Google Trends can be used to improve the predictions of economic indicators in Thailand's three sectors namely, (i) the labor market sector (unemployment rate and unemployment registration), (ii) the real sector (automobile sales), and (iii) the financial sector (SET index). These sectors were selected due to the following reasons. First, these sectors show strong evidence of activities moving towards online platforms. Second, they seem to be the sectors where search activities could, at least partially, translate into people behaviors. Finally, they are the sectors that the data can be easily accessed.

3.1 The Labor Market Sector

Thailand's labor market is composed of the formal sector and the informal sector. The formal sector includes those employed in private firms, governments, and state enterprises. The informal sector includes those employed in family businesses and those who are self-employed. Like many other developing countries, the majority of Thai workers are employed in the informal sector. Two interesting labor market indicators that will be examined under this section are (i) the unemployment rate and (ii) the unemployment registration (dismissed workers).

The unemployment rate, a monthly indicator administered by the National Statistical Office of Thailand (NSO), is calculated by dividing the number of unemployed workers by the number of those in the labor force. The unemployed workers are people who are not currently working (either in the formal sector or informal sector) but are looking for work or are available for work. The labor force is composed of the unemployed, the employed, and the people who are seasonally inactive. Thus, the unemployment rate reflects the unemployment situation for both formal and informal workers as a whole. It does not provide insights into the labor market situation of the formal sector and the informal sector separately.

The unemployment registration (dismissed workers), a monthly indicator collected by the Department of Employment (Ministry of Labor), contains the number of workers being dismissed from their formal sector jobs. Workers employed in the formal sector (excluding the public officials) are required to be insured under The Social Security Act B.E. 2533 (1990). With the social security, workers are eligible to receive unemployment benefits if they ever become unemployed. In the case of dismissal, the benefits are at the rate of 50% of their previous wage (for not more than 180 days). In the case of resignation, the benefits are at the rate of 30% of their previous wage (for not more than 90 days). However, in order to receive the unemployment benefits, workers will need to register their unemployment at the Department of Employment within 30 days of becoming unemployed. The Department of employment collected the unemployment registration separately for the dismissal case and the resignation case from July 2004 until May 2016.[8] In this study, we focus on the unemployment registration for the dismissal case (rather than the resignation case) since it appears to be a better proxy of the labor market situation in the formal sector.

For the labor market sector, there is some evidence that many activities are now being conducted online. Many job search websites have been launched in the past decades. One of the websites reported having more than 1.3 million resume postings and more than 80,000 job

---

[8] Although the unemployment registration and the unemployment benefit claim processes are still ongoing, unfortunately the Department of Employment no longer collects and manages the unemployment registration data (dismissed vs. resignation) and thus cannot make it available to the public.

postings currently listed.[9] Lekfuangfu, Nakavachara, and Sawaengsuksant (2017) reported that the number of online resume postings and online job postings in Thailand have been growing over time. Moreover, the newspaper career classified ads have now becoming less popular and some newspapers have been shut down due to technology disruption.[10] Although we acknowledge that online job searches may not be applicable for some sectors such as agriculture, it is still worthwhile to analyze the labor market sector using the online data.

We have tried potential keywords that may be entered by people who are looking for jobs or people who have been recently been dismissed from their jobs. These keywords (and the corresponding English translation) are shown in Table 1. The first column shows the correlations of these keywords with the monthly unemployment rate data from the NSO. The second column shows the correlations of these keywords with the monthly unemployment registration data (dismissed workers) from the Department of Employment. Since January 2004 is the earliest month in which the Google Trends data are available, we started our unemployment rate data series from then until May 2017. For unemployment registration data series, the data are made available only from July 2004 to May 2016. Therefore, that is the time period we conducted our analysis for that data series.[11] Among these potential keywords, the keyword "สมัครงาน" (Applying for jobs) has the highest correlation (0.7108) with the unemployment rate. And the keyword "ตกงาน" (Dismissed from jobs) has the highest correlation (0.6339) with the unemployment registration data. Therefore, we will use these two keywords for our empirical analyses. We contrasted the unemployment rate trend with Google Trends for the keyword "สมัครงาน" (Applying for jobs) in Figure 2 and we contrasted the

[9] www.jobthai.com (as of 22 August 2017)
[10] This issue is not restricted to just Thailand. New York Observer ended its print edition in 2016. Village Voice and TODAY Newspaper are ending their print edition in 2017. For Thailand, Banmuang newspaper ended its print edition at the end of 2016.
[11] Google Trends data are accessed during July-August 2017.

unemployment registration (dismissed workers) trend with Google Trends for the keyword "ตกงาน" (Dismissed from jobs) in Figure 3.

For our empirical analyses, the base model for both (i) the monthly unemployment rate and (ii) the monthly unemployment registration (dismissed workers) is the AR process as follow:[12]

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \varepsilon_{it} \quad (1)$$

y_t is the variable of interest, namely, (i) the monthly unemployment rate or (ii) the monthly unemployment registration (dismissed workers). *t* is the time variable which is month. y_{t-1} and y_{t-2} are the lag terms. ε_{it} is the error term. The time period for the unemployment rate model is from January 2004 to May 2017. The time period for the unemployment registration model is from July 2004 to May 2016.[13] The Dicky Fuller test was conducted and we could reject the Null of a unit root (non-stationary) process at 1% for the unemployment rate and at 5% for the unemployment registration (dismissed workers). Thus, the AR model can be used.

The model with Google Trends is as follow:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + b_3 G_t + \varepsilon_{it} \quad (2)$$

G_t is the monthly Google Trends for (i) "สมัครงาน" (Applying for jobs) for the unemployment rate model or (ii) "ตกงาน" (Dismissed from jobs) for the unemployment registration (dismissed workers) model. Robust standard errors are used in all of our models.

[12] Many variations of the AR models were estimated and the best-fitted model was selected.
[13] Google Trends data are accessed during July-August 2017.

To compare the forecast accuracy among the models, we examine different types of prediction errors namely, the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Mean Squared Error (MSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE).[14]

The regression results for the labor market sector indicators are shown in Table 2. Columns 1 and 2 display the results for the unemployment rate without and with the Google Trends variable, respectively. The lag variables are positive and significant under both specifications. The Google Trends variable is positive and significant at 1% under Column 2. Columns 3 and 4 display the results for the unemployment registration (dismissed workers) without and with the Google Trends variable, respectively. The Google Trends variable is positive and significant at 1% under Column 4. The Adjusted R-Squared is improved once the Google Trends variable is included in the model. In addition, the model with Google Trends also has better prediction accuracies under all measures of our interest, namely, AIC, BIC, MSE, MAE, and MAPE.

3.2 The Real Sector

The real sector of the economy is associated with the production of goods and services. Within the real sector, the automobile production and sales activities are being focused in this study. The automotive sector is one of the most important sectors in Thailand. As of 2016, Thailand is ranked number 12 as the world's largest motor vehicles (passenger cars and commercial vehicles) production base list, with the overall production of 1,944,417 vehicles.[15] Similar to the labor market sector, some activities within the real sector such as the automobile

---

[14] These are the measures commonly used in the literature. See Choi and Varian (2009a, 2009b, 2012) and McLaren and Shanbhogue (2011), for example.

[15] The International Organization of Motor Vehicle Manufacture (2016). Retrieved from http://www.oica.net/category/production-statistics/ (as of 3 August 2017).

sales appeared to have shifted towards the online platforms. Many official car dealers have established their websites so that customers can check the information and contact them electronically. There are quite a number of online communities in which people discuss and exchange information about car purchases. In addition, many used car online marketplaces have been launched in the past decades. One of these online marketplaces reported having more than 85,000 automobiles currently listed.[16]

Similar to the analyses conducted in the previous literature (for other countries), we hypothesize that the search volume for particular car brands may be correlated with the actual monthly sales of such car brands. The rationale behind this is that people usually search for information of the products they intend to buy. Although, our hypothesis here is that Google Trends is linked to the actual new automobile sales, we acknowledge that it is possible for Google Trends to pick up public interests towards automobile discussion forums or second-hand automobile market activities.[17] We calculate the correlations of the monthly sales volume of each car brand and the Google Trends using the car brands as keywords (both in English and in Thai). The monthly first-hand automobile[18] sales volume data (i.e., number of vehicles sold) by brands were retrieved from the CEIC database.[19] The data series is collected and updated monthly by Toyota Motor Thailand, Co. Ltd.[20] Since January 2004 is the earliest month in which the Google Trends data are available, we started our data series from then until May 2017.[21] Table 3 displays the correlations. It appears that the brands in the Thai language have

[16] rod.kaidee.com (as of 22 August 2017)

[17] For Thailand, the official statistics are only available for new car sales. There is no official statistics for used car sales. Therefore, for our empirical analysis, we only examined the market for new car sales.

[18] The term automobile here comprises of both passenger and commercial vehicles (pick-up cars included). Motorcycles are not included.

[19] CEIC database is a global database compiled and administered by CEIC Data Company, Ltd. The database includes updated economic data series on various sectors such as financial, banking, production, investment, etc.

[20] Toyota Motor Thailand, Co. Ltd compiles and updates the new automobile sales volume (i.e., number of vehicles sold) for all leading brands in Thailand.

[21] Google Trends data are accessed during July-August 2017.

higher correlations (with the actual sales volume) than the brands in the English language. Therefore, we will use the brands in Thai, namely, "โตโยต้า" (Toyota), "นิสสัน" (Nissan), "ฮอนด้า" (Honda), "มิตซูบิชิ" (Mitsubishi), or "มาสด้า" (Mazda), as the keywords for our empirical analyses. We show the trends from Google vs. the trends of the actual monthly sales volume of these car brands in Figures 4, 5, 6, 7, and 8.

We conducted the Dicky Fuller test for the sales volume data series of Toyota, Nissan, Honda, Mitsubishi, and Mazda and found that we could reject the Null of a unit root (non-stationary) process only for Toyota, Nissan, and Honda (at 1%, 1%, and 5%, respectively) but could not reject the Null for Mitsubishi and Mazda. Therefore, we could use the AR model only for Toyota, Nissan, and Honda. For Mitsubishi and Mazda, we tried the difference model and tested for unit root. It turned out we could reject the Null of a unit root (non-stationary) process at 1% for both data series. Therefore, we will use the difference model for Mitsubishi and Mazda.

In order to select our empirical model, we tried different variations of the AR models (with different and multiple lag terms). We then selected the model that is the best fit for the data. For Toyota and Nissan, the best-fitted model is the one that includes lag-1 and lag-12 of the sales volume variable. However, for Honda, the best-fitted model is the one that includes lag-1 and lag-2 of the sales volume variable.

Therefore, for our empirical analyses, the base model for Toyota and Nissan is the AR process as follow:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-12} + \varepsilon_{it} \tag{3}$$

Inserting Google Trends into the analysis, the model becomes:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-12} + b_3 G_t + \varepsilon_{it} \tag{4}$$

The base model for Honda is as follow:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \varepsilon_{it} \tag{5}$$

Inserting Google Trends into the analysis, the model becomes:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + b_3 G_t + \varepsilon_{it} \tag{6}$$

y_t is the variable of interest, namely, (i) the monthly sales volume for Toyota or (ii) the monthly sales volume for Nissan, and (iii) the monthly sales volume for Honda. *t* is the time variable which is month. y_{t-1}, y_{t-2} and y_{t-12} are the lag terms. ε_{it} is the error term. G_t is the monthly Google Trends for (i) "โตโยต้า" (Toyota), (ii) "นิสสัน" (Nissan), and (iii) "ฮอนด้า" (Honda). The time period for the analysis is from January 2004 to May 2017.[22] Robust standard errors are used in all models.

As already discussed, the sales volume data series of Mitsubishi and Mazda cars do not follow a stationary process. Therefore, we use the difference model as follow:[23]

$$\Delta y_t = a + b_1 \Delta y_{t-1} + b_2 \Delta y_{t-2} + \varepsilon_{it} \tag{7}$$

And the model with Google Trends is as follow:

$$\Delta y_t = a + b_1 \Delta y_{t-1} + b_2 \Delta y_{t-2} + b_3 G_t + \varepsilon_{it} \tag{8}$$

y_t is the variable of interest, namely, (i) the monthly sales volume for Mitsubishi or (ii) the monthly sales volume for Mazda. *t* is the time variable which is month. Δy_t is $y_t - y_{t-1}$; Δy_{t-1} is $y_{t-1} - y_{t-2}$; and Δy_{t-2} is $y_{t-2} - y_{t-3}$. ε_{it} is the error term. G_t is the monthly Google

[22] Google Trends data are accessed during July-August 2017.
[23] Many variations of the AR models were estimated and the best-fitted model was selected.

Trends for (i) "มิตซูบิชิ" (Mitsubishi), or (ii) "มาสด้า" (Mazda). The time period for the analysis is from January 2004 to May 2017.[24] Robust standard errors are used in all models.

To compare the forecast accuracy among the models, we examine different types of prediction errors namely, the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Mean Squared Error (MSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE).[25]

The regression results are shown in Table 4. The results from the AR models for Toyota (without and with the Google Trends), Nissan (without and with the Google Trends), and Honda (without and with the Google Trends) are shown in Columns 1, 2, 3, 4, 5, and 6. The lag variables are significant under all specifications. The Google Trends variable is positive and significant at 1%, 1%, and 5% for Toyota, Nissan, and Honda, respectively. For all of the three car brands, the Adjusted R-Squared is improved once the Google Trends variable is included in the model. In addition, the model with Google Trends also has better prediction accuracies under most measures of our interest, namely, AIC, MSE, MAE, and MAPE.

The results from the difference models for Mitsubishi (without and with Google Trends) and Mazda (without and with Google Trends) are shown in Columns 7, 8, 9, and 10, respectively. The lag difference variables are significant in all specifications. The Google Trends variable is positive and significant at 10% for Mitsubishi and 5% for Mazda. The Adjusted R-Squared is improved once the Google Trends variable is included in the models. In addition, the models with Google Trends also has better prediction accuracies under most measures of our interest, namely, AIC, MSE, and MAE.

---

[24] Google Trends data are accessed during July-August 2017.
[25] These are the measures commonly used in the literature. See Choi and Varian (2009a, 2009b, 2012) and McLaren and Shanbhogue (2011), for example.

3.3 The Financial Sector

The Stock Exchange of Thailand (SET) is the main stock market in Thailand. There are currently 592 companies registered under SET.[26]

The SET index includes all common stocks listed under SET and is calculated using the formula:

$$SET\ index = \frac{100 \times Market\ Value}{Base\ Market\ Value}$$

The Market Value represents the current market value of all stocks whereas the Base Market Value represents the market value of all stocks on 30 April 1975 (when SET was established).

Initially, prior to the era of online stock trading, investors needed to call their brokers to check the stock prices and to ask the brokers to execute the transactions for them. Thus, the investors could not retrieve the price information and could not execute the transactions on a real-time basis. In the year 2000, Settrade.com Co., Ltd. (Settrade), a subsidiary of SET, was established to develop the online stock trading platform and to provide the online trading service to investors. Currently, the investors can check the real-time stock price online and can also execute their transactions via an application called "Streaming" (developed by Settrade) via their smartphones. Thus, for the financial sector, it is obvious that the activities have been moving towards the online platform.

We have tried potential keywords that may be entered by people who are interested in stock trading and perhaps want to look for more information about the SET index. These

[26] There are 592 companies registered under SET and there are 139 companies registered under the Market for Alternative Investment (MAI), a sister market of SET for smaller market cap firms. Source: https://marketdata.set.or.th/mkt/sectorialindices.do (as of 3 August 2017)

keywords (and the corresponding English translation, when applicable) are shown in Table 5. The table also shows the correlations of these keywords with the SET index data.[27] Since January 2004 is earliest month that the Google Trends data are available, we started our data series from then until June 2017.[28] Among these potential keywords, the keyword "หุ้น" (Stock) has the highest correlation (0.9016) with the SET index. Therefore, we will use this keyword for our empirical analyses. We contrasted the monthly SET index data with Google Trends of the keyword "หุ้น" (Stock) in Figure 9.

We conducted the Dicky Fuller test for the SET index data and could not reject the Null of a unit root (non-stationary) process. Therefore, the standard AR model cannot be used. We then tried the difference model and conducted the test. It turned out we could reject the Null of a unit root (non-stationary) process at 1%. Therefore, we will use the difference model for the monthly SET index data.

The base model for the monthly SET index data is as follow:[29]

$$\Delta y_t = a + b_1 \Delta y_{t-1} + \varepsilon_{it} \tag{9}$$

And the model with Google Trends is as follow:

$$\Delta y_t = a + b_2 \Delta G_t + \varepsilon_{it} \tag{10}$$

y_t is the variable of interest, which is the monthly SET index data. *t* is the time variable which is month. Δy_t is $y_t - y_{t-1}$; and $\Delta \mathrm{y}_{t-1}$ is $\mathrm{y}_{t-1} - \mathrm{y}_{t-2}$. ε_{it} is the error term. G_t is the

[27] The SET index data were retrieved from https://www.set.or.th/en/market/market_statistics.html (as of 11 July 2017).
[28] Google Trends data are accessed during July-August 2017.
[29] Similar to the empirical analyses conducted in the previous sections, we tried many variations of the AR models and attempted to select the best-fitted model. However, for the SET index, it turned out that once we take the difference on the data, the lag terms no longer explain the data. But we showed that including the Google Trends variable can improve the model.

Google Trends for "หุ้น" (Stock). ΔG_t is $G_t - G_{t-1}$. The time period for the analysis is from January 2004 to June 2017.[30] Robust standard errors are used in all models.

To compare the forecast accuracy among the models, we examine different types of prediction errors namely, the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Mean Squared Error (MSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE).[31]

The regression results are shown in Table 6. Columns 1 and 2 display the results for the SET index without and with the Google Trends variable, respectively. The lag difference variable was not significant under Column 1. However, replacing Δy_{t-1} with ΔG_t improved the Adjusted R-Squared and the coefficient of ΔG_t is positive and significant. In addition, the model with Google Trends variable has better prediction accuracies under some measures of our interest, namely, MSE and MAE.

4. Conclusion

This paper illustrated how Google Trends can be used to improve the predictions of various economic indicators of Thailand. Specifically, the paper utilized the real-time aspect of Google Trends to conduct the nowcasting analyses – using the current period's real-time information to estimate the current period's economic indicators. The authors performed the nowcasting analyses in three sectors, namely, (i) the labor market sector, (ii) the real sector, and (iii) the financial sector.

The results revealed that, incorporating Google Trends data into the prediction models improved the Adjusted R-Squared and improved the prediction accuracies under various

[30] Google Trends data are accessed during July-August 2017.
[31] These are the measures commonly used in the literature. See Choi and Varian (2009a, 2009b, 2012) and McLaren and Shanbhogue (2011), for example.

measures. Our results are in line with Choi and Varian (2009a, 2009b, 2012) that have tried to use Google Trends to nowcast similar economic indicators in the United States. Our results are also in line with other literature that conducted the analyses for advanced countries such as the United Kingdom (McLaren and Shanbhogue (2011)), Germany (Askitas and Zimmermann (2009)), and France (Fonduer and Karame (2013)) and for emerging middle income countries such as Chile (Carriere-Swallow and Labbe (2013), Turkey (Chadwick and Sengul (2012) and Zeybek and Ugurlu (2015)), and Central American countries (Seabold and Coppola (2015)).

The propose of this study is neither to convince the readers that Google Trends data are flawless nor to affirm that we could rely completely on the Google Trends data for nowcasting. Obviously, there are still some sectors that Google Trends data are not applicable, for example, the agricultural sector and other sectors that the majority of the people are not internet users. Moreover, the correlations between Google Trends keywords and the actual economic indicators are sometimes noisy. In addition, the fact that Google does not reveal the exact methodology that it uses to calculate the Google search volume index, makes it hard for researchers to draw powerful conclusions out of the analyses that utilized the Google Trends data.

However, what this study tries to argue is that, despite the drawbacks, the information retrieved from Google searches can still be shown useful in many cases. For Thailand, the Google Trends data were proven useful in nowcasting various economic indicators in three sectors, namely, (i) the labor market sector, (ii) the real sector, and (iii) the financial sector. As already mentioned, Google is currently the most-used search engine in the world and there are approximately 3.5 billion searches being conducted on Google each day. Therefore, the search data collected by Google are too important to be ignored.

In the future, economics research will be driven more and more by data. In the age of digital economy, the new major source of data for research is data from the internet, like Google Trends and many others. The authors hope to see many more movements towards the idea of open data (of course, with appropriate measures being taken so that personal/sensitive information is protected). With open data, various researchers can fully utilize the data and help modify the existing methodology currently being applied to the data. Perhaps, the shortcomings of the data can be fixed and the efficiency of how the data are processed can be improved. Under this environment, many more meaningful research questions can be asked and many more rigorous analyses can be conducted.

5. References

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), 107-120.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1), 68-105.

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Chadwick, M. G. & Sengul, G. (2012). Nowcasting Unemployment Rate in Turkey: Let's Ask Google. Central Bank of the Republic of Turkey. Working Paper No. 12/18.

Choi, H., & Varian, H. (2009a). Predicting the present with Google Trends. *Google Inc*, 1-20.

Choi, H., & Varian, H. (2009b). Predicting initial claims for unemployment benefits. *Google Inc*, 1-5.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2-9.

Fonduer, Y., & Karame, F. (2013). Can Google Data Help Predict French Youth Unemployment? *Economic Modelling*, 30, 117-125.

Lekfuangfu, N.W., Nakavachara, V., & Sawaengsuksant, P. (2016). Labour Market Insights: the Power of Internet-Based Data. Conference Paper presented at the Bank of Thailand's Symposium 2016.

Lekfuangfu, N.W., Nakavachara, V., & Sawaengsuksant, P. (2017). Glancing at Labour Market Mismatch with User-Generated Internet Data. PIER Discussion Paper No. 53.

McLaren, N. & Shanbhogue, R. (2011). Using Internet Search Data as Economic Indicators. Bank of England Quarterly Bulletin, 2011Q2, 134-140.

Seabold, S., & Coppola, A. (2015). Nowcasting Prices Using Google Trends: An Application to Central America. World Bank Group Policy Research Working Paper 7398.

Suhoy, T. (2009). Query Indices and a 2008 Downturn: Israeli Data. Bank of Israel Discussion Paper No. 2009.06.

Vincente, M. R., Lopez-Menendez, A. J., & Perez, R. (2015). Forecasting Unemployment with Internet Search Data: Does It Help to Improve Predictions when Job Destruction is Skyrocketing? *Technological Forecasting & Social Change*, 92, 132-139.

Zeybek, O. & Ugurlu, E. (2015). Nowcasting Credit Demand in Turkey with Google Trends Data. *Journal of Applied Economic Science*, Vol. 10, Issue 2(32), 293-300.

Appendix

Figure 1: Google Trends for the Keyword "หวย" (Informal Thai word for "Lottery")

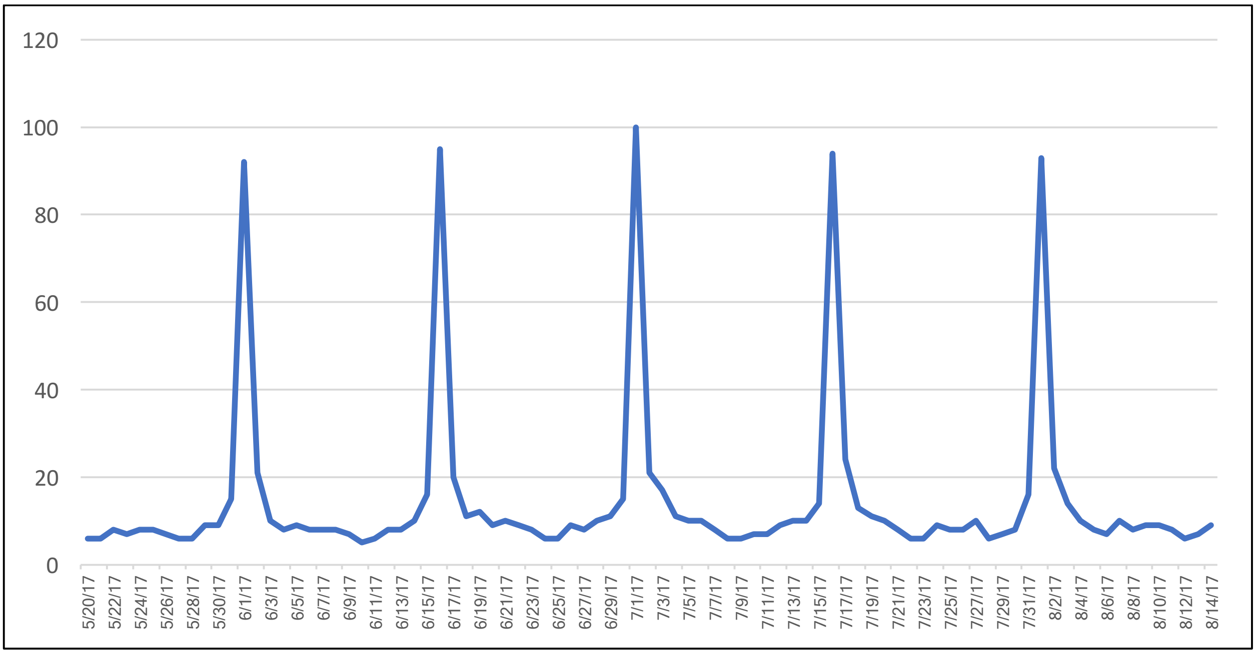

Figure 2: Unemployment Rate (%) and Google Trends

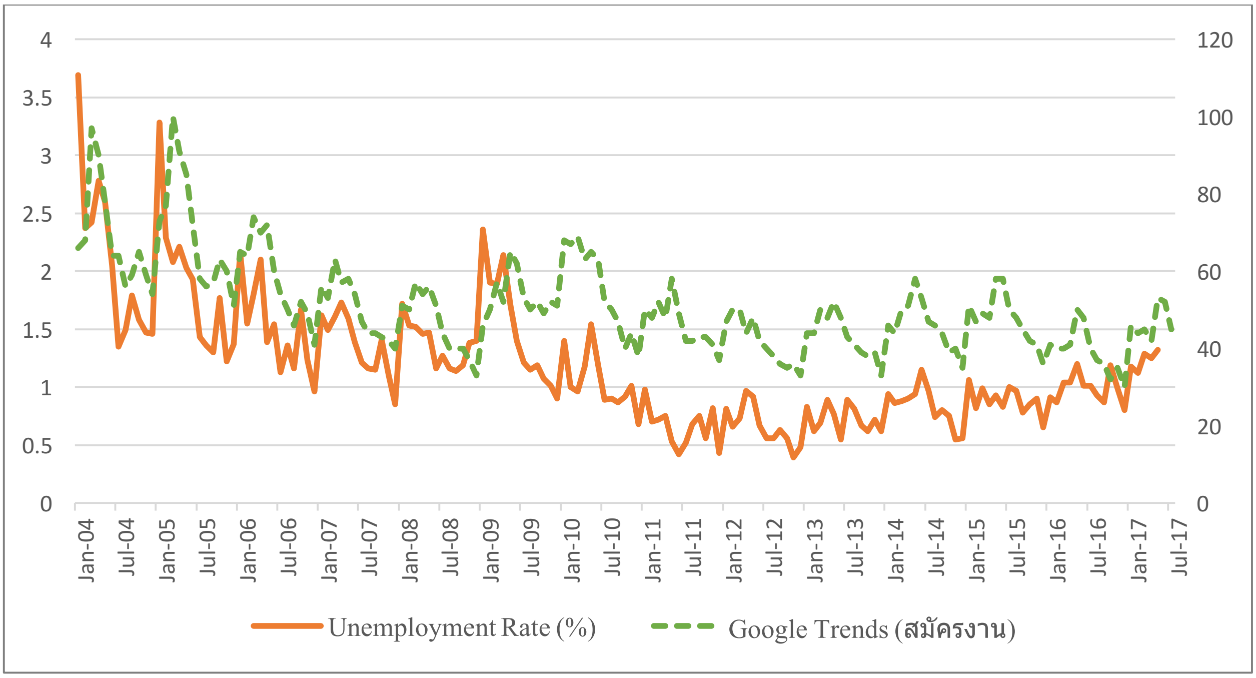

Figure 3: Registration for Unemployment (Dismissed Workers) and Google Trends

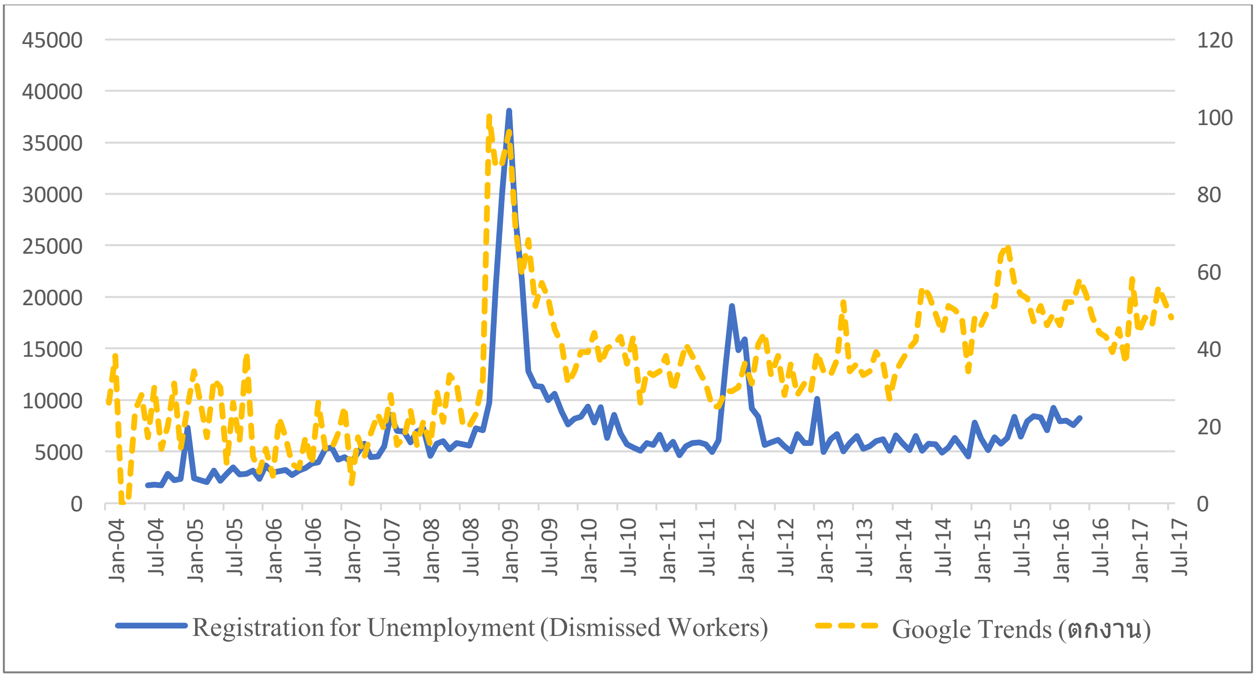

Figure 4: Automobile Sales (Toyota) and Google Trends

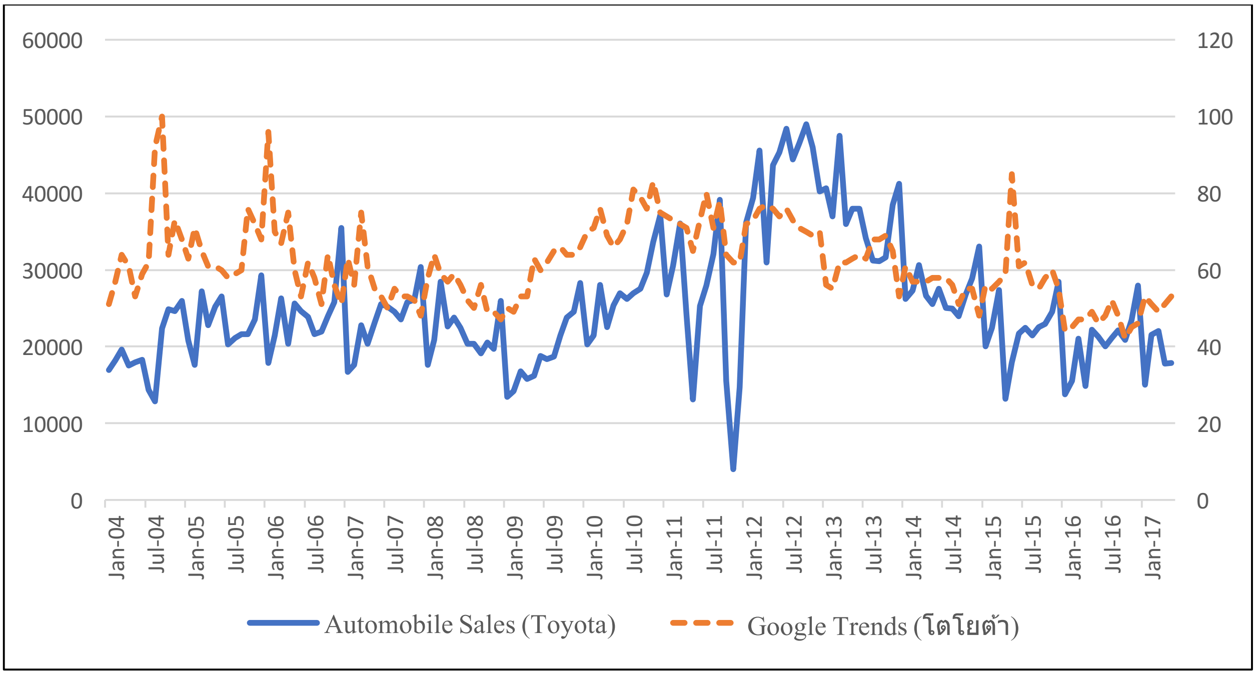

Figure 5: Automobile Sales (Nissan) and Google Trends

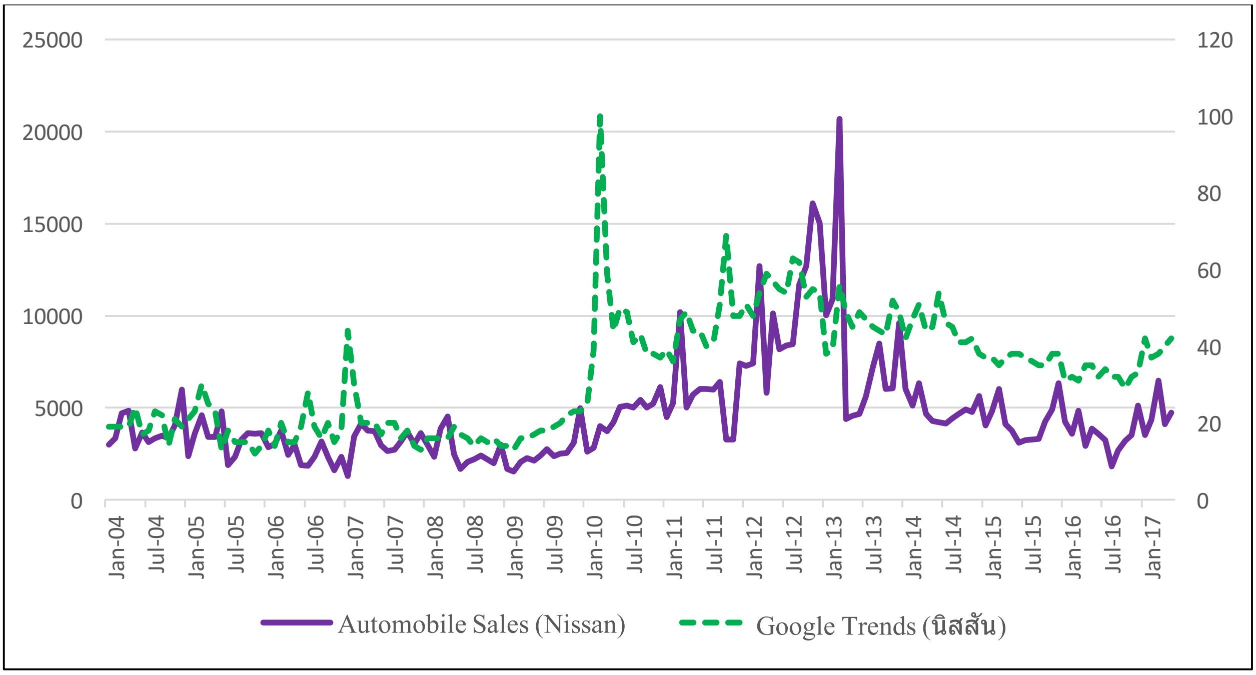

Figure 6: Automobile Sales (Honda) and Google Trends

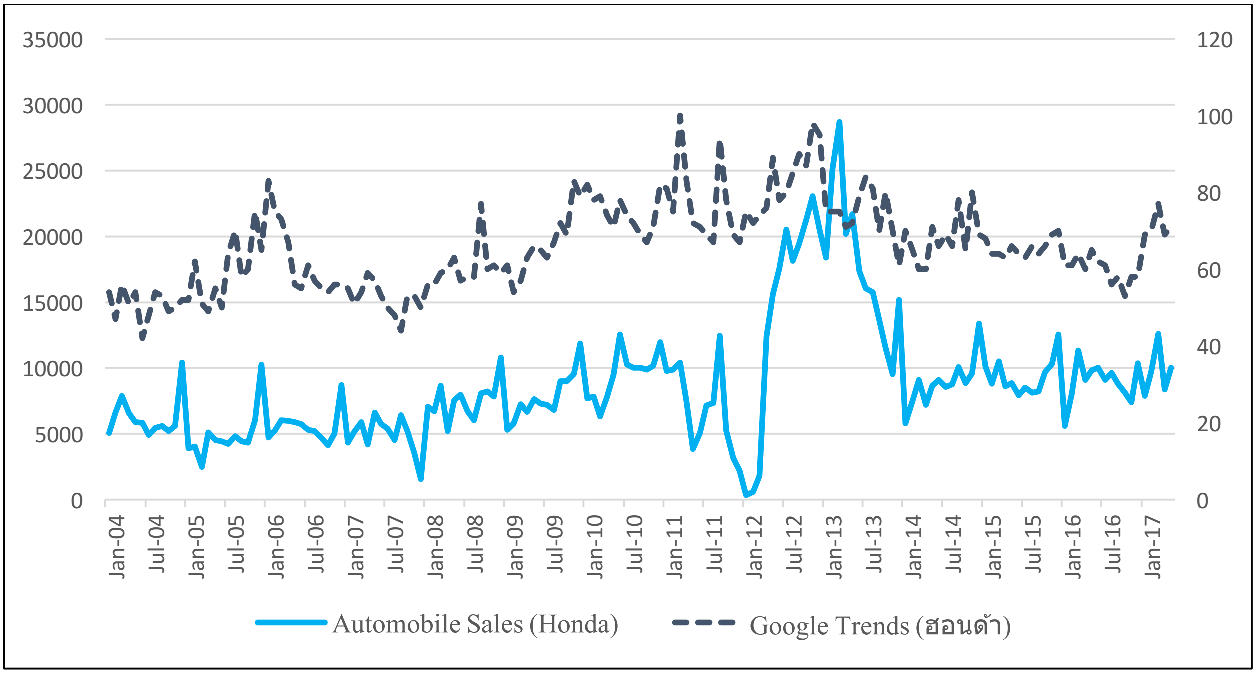

Figure 7: Automobile Sales (Mitsubishi) and Google Trends

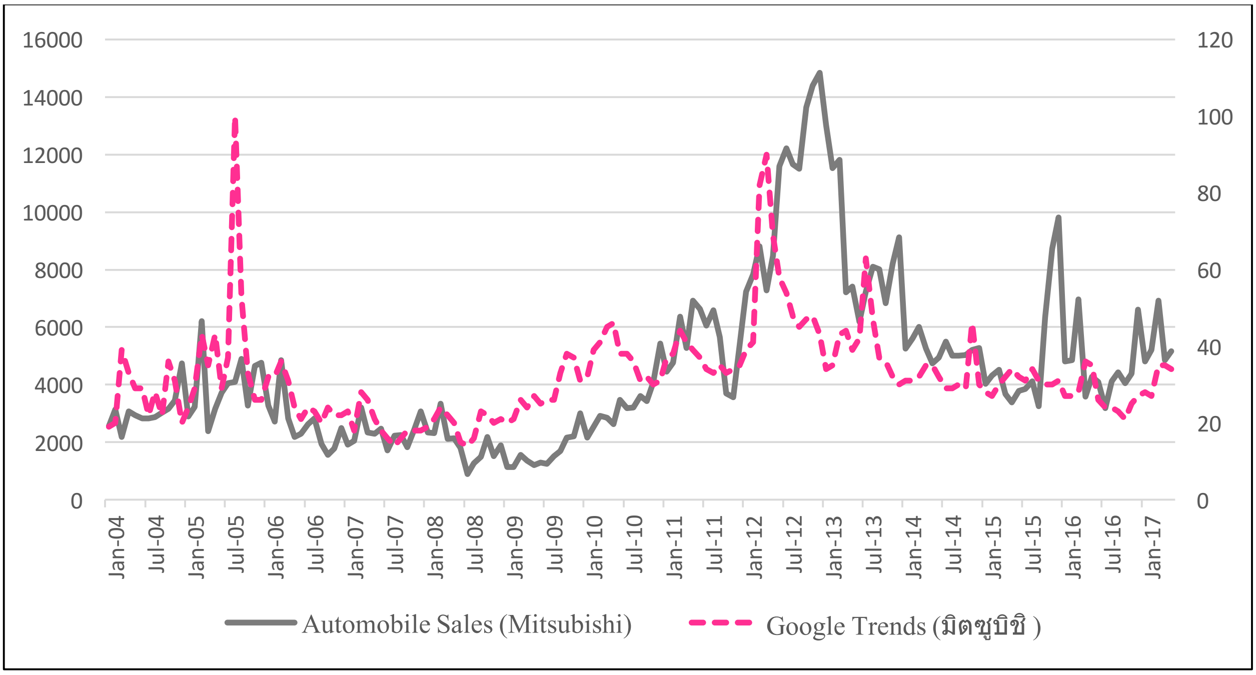

Figure 8: Automobile Sales (Mazda) and Google Trends

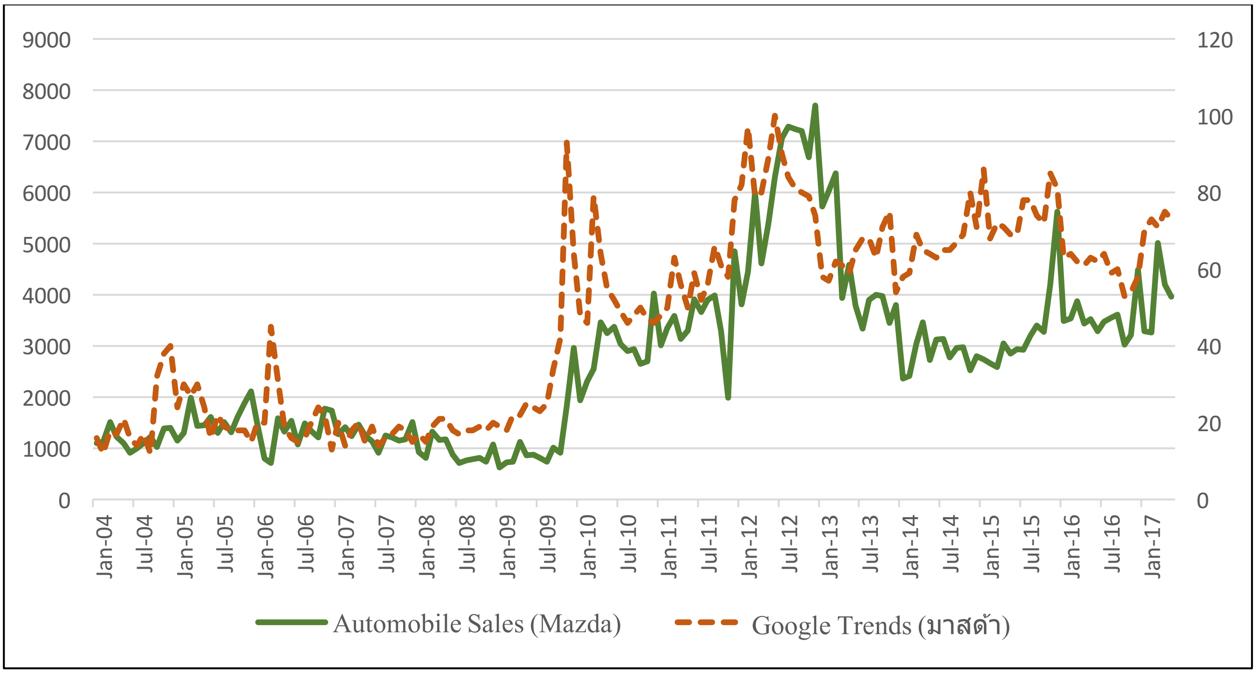

Figure 9: SET Index and Google Trends

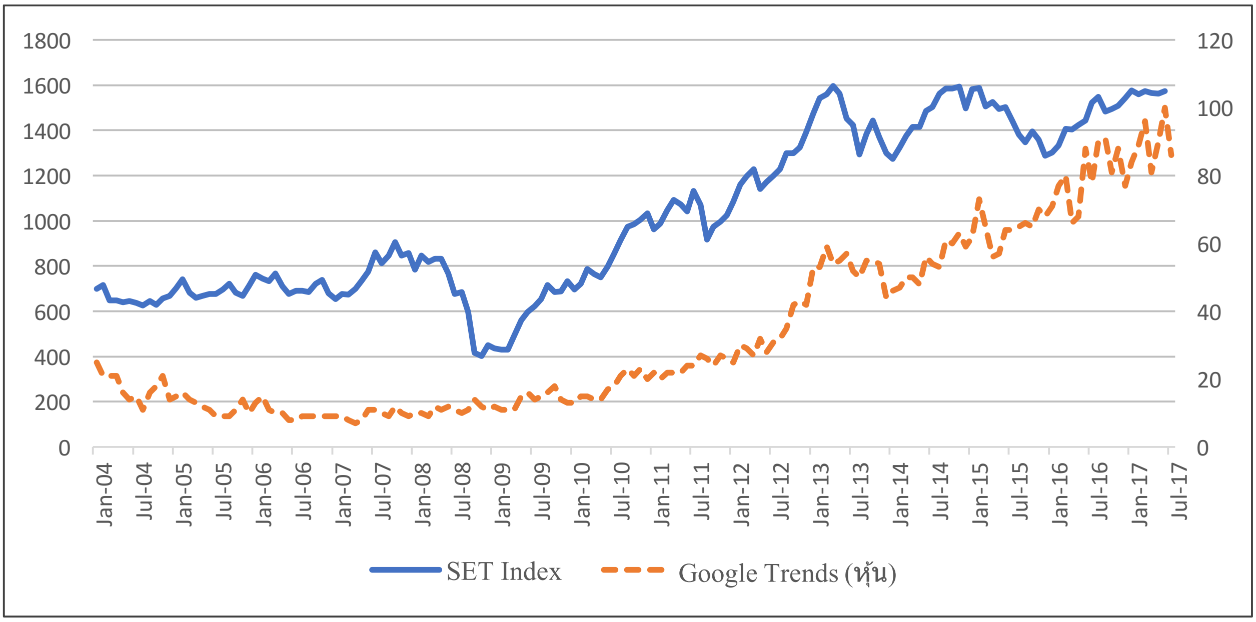

Table 1: Potential Keywords (Labor Market Sector)

Potential Keywords*	(1) Correlations Unemp	(2) Correlations SSLaidOff
สมัครงาน (Applying for Jobs)	**0.7108**	-0.2041
หางาน (Searching for Jobs)	0.645	-0.0387
ตกงาน (Dismissed from Jobs)	-0.1857	**0.6339**
ว่างงาน (Unemployed)	0.3277	0.1882
ประกันสังคมว่างงาน (Social Security for Unemployment)	-0.0044	0.4954
ประกันสังคม (Social Security)	0.3479	0.5419
เงินทดแทน (Severance Pay)	0.6212	0.0164

*English translation in parentheses

Table 2: Regression Results (Labor Market Sector)

VARIABLES	(1) Unemp	(2) Unemp	(3) SSLaidOff	(4) SSLaidOff
L.Unemp	0.5817***	0.4865***		
	(0.0567)	(0.0672)		
L2.Unemp	0.2438***	0.1419**		
	(0.0518)	(0.0694)		
GG_Apply		0.0106***		
		(0.0039)		
L.SSLaidOff			1.0822***	0.9675***
			(0.2186)	(0.1855)
L2.SSLaidOff			-0.2492	-0.2525*
			(0.1555)	(0.1435)
GG_Dismissed				60.0971***
				(19.5066)
Constant	0.1946***	-0.1100	1,189.5283**	-120.8942
	(0.0480)	(0.1222)	(575.0500)	(647.9553)
Observations	159	159	141	141
R-squared	0.6756	0.7056	0.7688	0.7986
Model	Level	Level	Level	Level
Period	1/2004-5/2017	1/2004-5/2017	7/2004-5/2016	7/2004-5/2016
Adj R-Squared	0.671	0.700	0.765	0.794
AIC	57.80	44.40	2597	2579
BIC	67	56.70	2605	2591
MSE	0.08111	0.07362	5581142	4862222
MAE	0.1881	0.1845	1496	1473
MAPE	0.1727	0.1695	0.2266	0.2233

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Potential Keywords (Real Sector: Automobile Sales)

Potential Keywords*	(1) Correlations Toyota Sales	(2) Correlations Nissan Sales	(3) Correlations Honda Sales	(4) Correlations Mitsubishi Sales	(5) Correlations Mazda Sales
Toyota	-0.0992				
โตโยต้า (Toyota)	**0.2955**				
Nissan		0.279			
นิสสัน (Nissan)		**0.5885**			
Honda			-0.0113		
ฮอนด้า (Honda)			**0.5646**		
Mitsubishi				-0.0981	
มิตซูบิชิ (Mitsubishi)				**0.5254**	
Mazda					0.3201
มาสด้า (Mazda)					**0.8105**

*English translation in parentheses (if applicable)

Table 4: Regression Results (Real Sector: Automobile Sales)

VARIABLES	(1) Toyota	(2) Toyota	(3) Nissan	(4) Nissan	(5) Honda	(6) Honda	(7) D.Mitsubishi	(8) D.Mitsubishi	(9) D.Mazda	(10) D.Mazda
L.Toyota	0.6611***	0.5967***								
	(0.0792)	(0.0848)								
L12.Toyota	0.1972***	0.2222***								
	(0.0647)	(0.0662)								
GG_Toyota		148.3361***								
		(56.6567)								
L.Nissan			0.5933***	0.4885***						
			(0.1754)	(0.1859)						
L12.Nissan			0.2721**	0.2139**						
			(0.1049)	(0.1000)						
GG_Nissan				42.9338***						
				(16.2122)						
L.Honda					0.6841***	0.6186***				
					(0.1188)	(0.1211)				
L2.Honda					0.1888*	0.1740*				
					(0.1042)	(0.1018)				
GG_Honda						60.7730**				
						(24.6411)				
LD.Mitsubishi							-0.4910***	-0.5251***		
							(0.1749)	(0.1822)		
LD2.Mitsubishi							0.2475**	0.2663**		
							(0.1060)	(0.1116)		
GG_Mitsubishi								13.7916*		
								(8.1133)		
LD.Mazda									-0.7611***	-0.7908***
									(0.1868)	(0.1837)
LD2.Mazda									0.2966***	0.3097***
									(0.1086)	(0.1055)
GG_Mazda										4.3445**
										(1.8024)
Constant	3,644.8556**	-4,437.0838	622.9072	-60.9767	1,132.6725**	-2,185.9643*	28.2175	-417.1395*	31.9135	-172.3460***
	(1,808.4738)	(3,284.2419)	(597.5886)	(482.7816)	(539.6347)	(1,180.0794)	(92.5633)	(251.6190)	(46.3127)	(65.3231)
Observations	149	149	149	149	159	159	158	158	158	158
R-squared	0.5434	0.5715	0.5490	0.5816	0.7229	0.7377	0.0964	0.1155	0.2036	0.2311
Model	Level	Level	Level	Level	Level	Level	Diff	Diff	Diff	Diff
Period	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017
Adj R-Squared	0.537	0.563	0.543	0.573	0.719	0.733	0.0847	0.0983	0.193	0.216
AIC	2998	2991	2683	2674	2941	2934	2684	2683	2461	2458
BIC	3007	3003	2692	2686	2950	2947	2693	2695	2470	2470
MSE	3.08E+07	2.89E+07	3718057	3448814	6088962	5763856	1344227	1315699	328522	317174
MAE	3929	3762	1169	1136	1755	1670	800.7	787.8	378	376.6
MAPE	0.1884	0.1823	0.2594	0.2496	0.3047	0.3223	1.41	1.5	1.444	1.78

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Potential Keywords (Financial Sector)

Potential Keywords*	Correlations SET Index
หุ้น (Stock)	**0.9016**
ราคาหุ้น (Stock Price)	0.8998
ตลาดหุ้น (Stock Market)	0.752
SET	0.0031
SET Index	0.6816

*English translation in parentheses (if applicable)

Table 6: Regression Results (Financial Sector)

VARIABLES	(1) D.SET	(2) D.SET
LD.SET	0.1281	
	(0.0937)	
D.GG_Stock		1.4681**
		(0.7010)
Constant	4.6747	4.7561
	(4.0377)	(3.8858)
Observations	160	161
R-squared	0.0164	0.0172
Model	Diff	Diff
Period	1/2004-6/2017	1/2004-6/2017
Adj R-Squared	0.0102	0.0110
AIC	1702	1712
BIC	1708	1718
MSE	2383	2380
MAE	38.18	37.91
MAPE	3.704	3.812

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1