Thailand and The World Economy

P-ISSN: 2630-0931
E-ISSN: 2651-0529

Homepage : https://www.tci-thaijo.org/index.php/TER

Vol. 36, No. 3, September - December 2018 Page [23-46]

Predicting the Present Revisited: The Case of Thailand

*Voraprapa Nakavachara**
Faculty of Economics, Chulalongkorn University, Thailand
Nuarpear Warn Lekfuangfu
Faculty of Economics, Chulalongkorn University, Thailand
Centre for Economic Performance, London School of Economics and Political Science, United Kingdom

Received 20 September 2017, Received in revised form 5 March 2018,
Accepted 22 March 2018, Available online 24 December 2018

Abstract

Google is currently the most widely used search engine in the world. There are approximately 3.5 billion searches 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 current period real-time information to estimate current period indicators of interest. In this paper, we show how Google Trends can be used for nowcasting various Thai economic indicators. The areas analyzed are (i) the labor market sector (unemployment registration and unemployment rate), (ii) the real sector (automobile sales), and (iii) the financial sector (the SET index). The results revealed that incorporating Google Trends data into prediction models improved both the Adjusted R-Squared and predication accuracies under various measures.

Keywords: Nowcasting, Google Trends

JEL Classification: J01, L62, G10, G17

***Corresponding author:** Address: Phayathai Road, Pathumwan, Bangkok 10330, Thailand. Email: voraprapa.n@chula.ac.th

1. Introduction

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

Scholars have tried to study and utilize the search information 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 rates and initial claims for unemployment insurance.

One of the most important advantages inherent in 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 Google Trends data. Thus, the real-time aspect of Google Trends 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 entails using the previous period's data to predict the following period's economic indicators, nowcasting simply means using current period data to predict current period economic indicators.

The literature has documented successful attempts in using Google Trends to improve the predictions of many economic indicators in various countries. Since current-period Google Trends information can be retrieved almost in real-time, much earlier than the time at which current period economic indicators have traditionally become available, incorporating Google Trends information into the nowcasting model can improve predictions. Choi and Varian (2009a, 2009b, 2012), Askitas and Zimmermann (2009), Suhoy (2009), McLaren and Shanbhogue (2011), Carriere-Swallow and Labbe (2013), Fonduer and Karame (2013), Vincente, Lopez-Menendez, and Perez (2015) and Seabold and Coppola (2015) have all produced studies illustrating such prediction methodologies and how Google Trends improved outcomes in the context of their research projects.

However, nowcasting using Google Trends does have drawbacks. Google search volume indices, although revealing public interests at the time of the search involved, 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

---

[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)

information on the subject. It does not reveal their opinions about the subject. Thus, the correlations of Google Trends with actual economic indicators could be noisy. In addition, Google does not reveal the exact methodology that it uses in calculating the Google search volume indices. Thus, researchers can never cross-check calculations and will have to rely on Google to ensure the accuracy and the consistency of the data concerned. Despite such drawbacks, Google Trends still provides useful real-time information and, thus, potentially improves predictions concerning many economic indicators.

As already discussed, from 2016 to 2017, the number of internet users grew 10% globally. The growth rate was highest, at 15%, for the Asia-Pacific region.[4] Within the Asia-Pacific region, Thailand is a country with a particularly high internet penetration rate of 67%.[5] In 2017, the country had 46 million internet users, a 21% increase from the previous year. There were 47.9 million unique mobile device users. Approximately 11.58 million people reported having conducted an online purchase and the country's total revenue derived from the e-commerce market (in 2016) stood at USD 2.8 billion.[6] With such a significant volume of internet activity, Thailand makes an interesting case study. Within emerging middle-income countries, only a few studies have explored the potential of Google Trends in predicting economic indicators. Carriere-Swallow and Labbe (2013) studied the role of Google Trends in nowcasting the automobile market in Chile. Regarding Turkey, Chadwick and Sengul (2012) studied how Google Trends could help predict the country's unemployment rate and Zeybek and Ugurlu (2015) studied how Google Trends helped predict the country's credit demand. Seabold and Coppola (2015) explored how Google Trends was able to help predict price levels in Costa Rica, El Salvador and Honduras. However, within Thailand, to the best of our knowledge, besides an earlier version of this paper (Lekfuangfu, Nakavachara, and Sawaengsuksant (2016)), there is currently no research paper investigating how Google Trends is able to improve predictions concerning 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 on three areas, namely, (i) the labor market sector, (ii) the real sector, and (iii) the financial sector. The paper is organized as follows. Section 2 discusses Google Trends and how it has emerged over time. Section 3 uses Google Trends data to nowcast several of Thailand's economic indicators. The econometric models and results obtained are discussed under this section. Section 4 concludes the paper and discusses the authors' viewpoint regarding the future of economic research under an open data environment.

2. Evolution of Google Trends

Google Trends, first launched in 2004, is a web service provided by Google that reports trends of search keywords being conducted on Google's search engine. Specifically, Google Trends reports search volume indices – indicating how often keywords have been searched relative to the total number of searches at the same time/location. A particular index is normalized to be in the range of 0 to 100 over the

---

[4] Tied 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)

selected time period. One can retrieve search volume index data dated back to January 2004.[7]

Initially, Google Trends data could be retrieved on a weekly basis dating back to January 2004. (Other frequency types could be retrieved but with shorter time spans.) In 2009, Hal Varian, the chief economist of Google, first wrote papers on how Google Trends could 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 property of nowcasting using Google Trends is that data becomes available more frequently and sooner than actual official economic indicators. Therefore, current period Google Trends data is usually already available and able to be used in the prediction of current-period economic indicators (which previously came out later after a period had ended). Subsequently, many studies (see the previous section) have followed methodologies using weekly Google Trends data to nowcast monthly economic indicators.

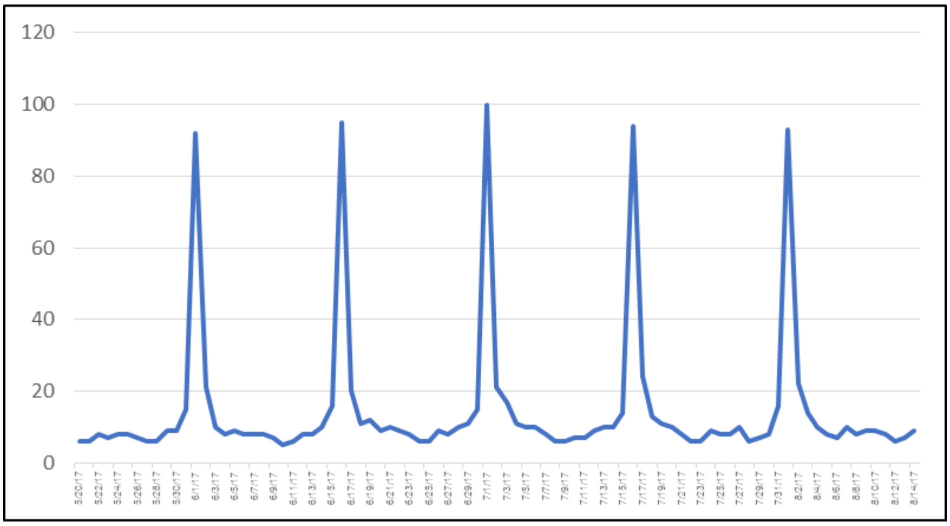
As of mid 2016, Google changed how Google Trends data was released to the public. The default frequency type that Google releases to the public when one retrieves data was dated back to January 2004 and is monthly instead of weekly. (Other frequency types can be retrieved, but with shorter time spans.)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

With a 67% internet penetration rate, Thailand represents an interesting case study on how search-generated data such as Google Trends can reflect public interests that might potentially translate into peoples' actual behavior. 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 a 90-day period in which daily data can be retrieved, restricting the location of the search to Thailand. Figure 1 provides supportive evidence for the coexistence of people's online search behavior and their real-world activities from two standpoints. First, the co-movement of both trendlines shows that people's online search behavior corresponds to 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 search engines, especially Google, is not restricted to only more sophisticated and educated groups within Thai society. Purchasing of state lottery tickets, exceedingly popular with approximately 71 million tickets being issued each round, is highly concentrated among members of the lower socio-economic class. Therefore, the behavior captured in Figure 1 suggests that online search engines are widely used across the whole social spectrum in Thailand.

---

[7] The history of the internet may be traced back to the 1960s when computers were connected for the first time and the first message were sent between them. However, it was not until the early 1990s when the internet was made available to the 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 widely used search engine in the world with approximately 3.5 billion searches being conducted each day.

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



Source: Retrieved from https://trends.google.co.th/trends/?geo=TH with the keywords

3. Nowcasting Thailand

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

3.1 The Labor Market Sector

Thailand's labor market is composed of formal and informal sectors. The formal sector includes workers employed in private firms, government organizations and state enterprises. The informal sector includes workers employed in family businesses and the self-employed. Like many 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) unemployment registration (dismissed workers) and (ii) the unemployment rate.

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

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 total number of those in the labor force. Unemployed workers refer to people who are not currently working (either in the formal or informal sectors), but are looking for or 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 insight into the labor market situation of the formal sector and the informal sector separately.

Table 1 provides summary statistics of the labor market variables used in this study. The unemployment registration (dismissed workers) ranged from a minimum of 1,722 to 38,103 per month with an average of 6,884 per month. The unemployment rate ranged from 0.39% to 3.69% with an average of 1.19%.

Regarding 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 website reported having more than 1.3 million resume postings and more than 80,000 job postings currently listed.[9] Lekfuangfu, Nakavachara, and Sawaengsuksant (2017) reported that the number of online resume and job postings in Thailand has been growing exponentially over time. Moreover, newspaper career classified ads are becoming less popular with many newspapers ceasing print editions altogether.[10] Although we acknowledge that online job searches may not be applicable for some sectors such as agriculture, we believe that it is still worthwhile to analyze the labor market sector using such online data

Table 1: Summary Statistics of the Variables (Labor Market Sector)

Variable Name	Variable Description	Duration	Min	Max	Avera ge
SSLaidOff	Unemployment Registration (Dismissed Workers)	Jul 2004 - May 2016	1,72 2	38,1 03	6,884
Unemp	Unemployment Rate	Jan 2004 - May 2017	0.39	3.69	1.19

Source: Department of Employment, Ministry of Labor.

In this study we identified potential keywords that may be entered by people either looking for jobs or who have recently been dismissed from their jobs. These keywords (and their corresponding English translation) are shown in Table 2. The first column shows correlations of these keywords with monthly unemployment registration data (dismissed workers) from the Department of Employment. The second column

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

[9] www.jobthai.com (as of 22 August 2017)

[10] This issue is not restricted to just Thailand. The New York Observer ended its print edition in 2016 and the Village Voice and TODAY Newspaper are ending their print editions in 2017. In Thailand the Banmuang newspaper ended its print edition at the end of 2016.

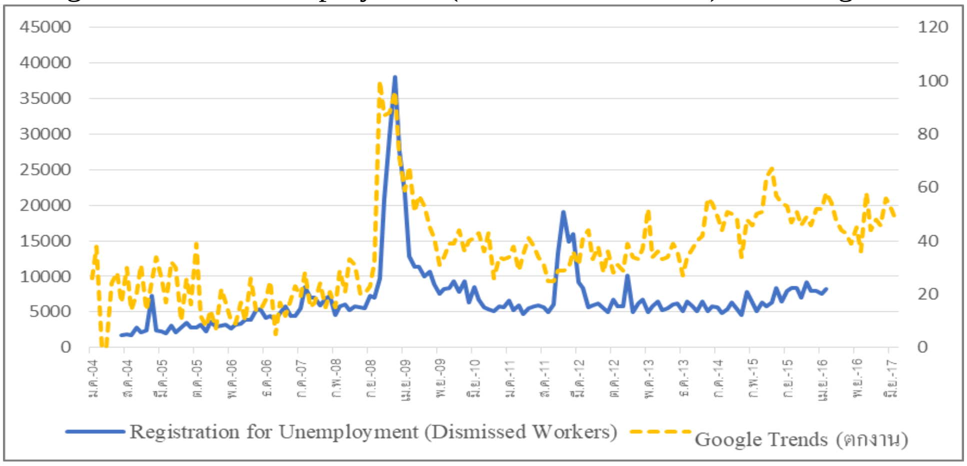
shows correlations of keywords with monthly unemployment rate data from the NSO. Since January 2004 is the earliest month in which Google Trends data is available, we started our unemployment rate data series from then through to May 2017. In terms of our unemployment registration data series, data was available only for July 2004 to May 2016. Therefore, that is the time period for which we conducted our initial analysis for that data series.[11] Among the potential keywords, the keyword "ตกงาน" (dismissed from job) has the highest correlation (0.6339) with unemployment registration data. The keyword "สมัครงาน" (applying for job) has the highest correlation (0.7108) with the unemployment rate. Therefore, we will use these two keywords in our empirical analyses. We contrasted the unemployment registration (dismissed workers) trend with Google Trends for the keyword "ตกงาน" (dismissed from job) in Figure 2 and we contrasted the unemployment rate trend with Google Trends for the keyword "สมัครงาน" (applying for job) in Figure 3.

Table 2: Potential Keywords (Labor Market Sector)

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

Note:*English translation in parentheses
Source: Department of Employment, Ministry of Labor, and retrieved Google Trends from https://trends.google.co.th/trends/?geo=TH with the keywords.

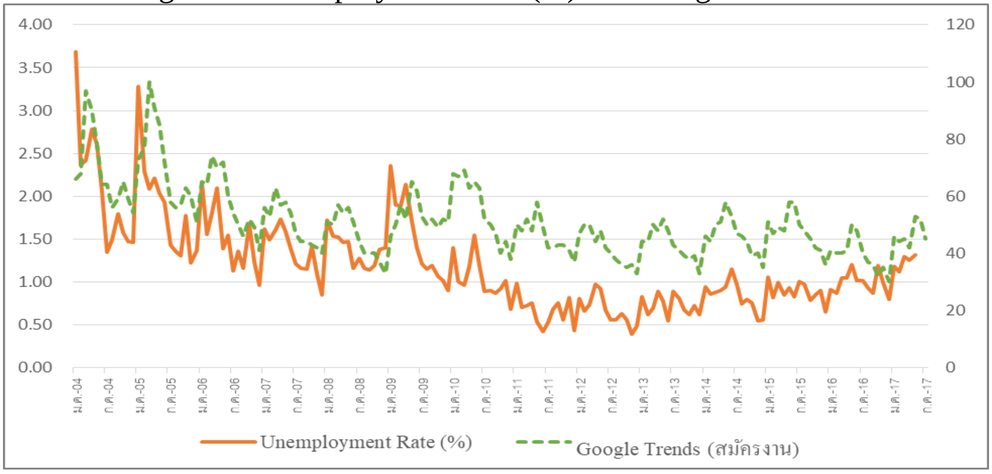
Figure 2: Registration for Unemployment (Dismissed Workers) and Google Trends



Source: Department of Employment, Ministry of Labor, and retrieved Google Trend from https://trends.google.co.th/trends/?geo=TH with the keywords.

[11] Google Trends data was accessed during July-August 2017.

Figure 3: Unemployment Rate (%) and Google Trends



Source: Department of Employment, Ministry of Labor, and retrieved Google Trends from https://trends.google.co.th/trends/?geo=TH with the keywords.

In our empirical analyses, the base model for both (i) monthly unemployment registration (dismissed workers) and (ii) the monthly unemployment rate follow the SARIMA process as follows:

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

y_t is the variable of interest, namely, (i) the natural log of the monthly unemployment registration (dismissed workers) or (ii) the natural log of the monthly unemployment rate. t is the time variable which is month. $\Delta_{12}y_t$ is $y_t - y_{t-12}$; and $\Delta_{12}\mathrm{y}_{t-1}$ is $\mathrm{y}_{t-1} - \mathrm{y}_{t-13}$. ε_{it} is the error term. The term $\Delta_{12}y_t$ (year-on-year change of the variable) is used in the model to mitigate for seasonality effects and non-stationary issues that may occur within the data series.

Since there could be other external factors that affect unemployment registration (dismissed workers) and/or the unemployment rate, we also estimate the model with additional explanatory variables as follows:

$$\Delta_{12}y_t = a + b_1\Delta_{12}y_{t-1} + b_2\Delta_{12}x_t + \varepsilon_{it} \tag{2}$$

x_t is the vector of additional explanatory variables, namely, (i) the natural log of the Consumer Price Index (CPI), (ii) the natural log of the policy interest rate, (iii) the natural log of the agricultural production index, and (iv) the natural log of the Manufacturing Production Index (MPI). $\Delta_{12}\mathrm{x}_t$ is $\mathrm{x}_t - \mathrm{x}_{t-12}$ or the year-on-year change of each of the variables. The relationship between unemployment rate and inflation (and sometimes the interest rate) has been widely discuss in the literature (see Phillips, A. W. (1958), for example). In addition, Thai labor market conditions rely heavily on the agricultural sector and the manufacturing sector. Therefore, these are the variables that we choose to include in our models.

We obtained the CPI data from the Bureau of Trade and Economic Indices, Ministry of Commerce and the policy interest rate data from the Bank of Thailand, while we sourced agricultural production indices from the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives. Note that since the earliest data available for this agricultural production index is January 2005, the time period for the empirical analysis conducted under this section started from January 2005 (i.e., January

2005 to May 2016 for the unemployment registration model and January 2005 to May 2017 for the unemployment rate model). Finally, we obtained MPI data from the Bank of Thailand and the Office of Industrial Economics, Ministry of Industry.[12]

The model for Google Trends is as follows:

$$\Delta_{12}y_t = a + b_1\Delta_{12}y_{t-1} + b_3\Delta_{12}G_t + \varepsilon_{it} \tag{3}$$

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

G_t is the natural log of monthly Google Trends for (i) "ตกงาน" (dismissed from job) for the unemployment registration (dismissed workers) model and (ii) "สมัครงาน" (applying for job) for the unemployment rate model. $\Delta_{12}\mathrm{G}_t$ is $\mathrm{G}_t - \mathrm{G}_{t-12}$ or the year-on-year change of the variable. Robust standard errors are used in all of our 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), and the Out-of-sample Mean Squared Error (Out-of-sample MSE).[13] The Out-of-sample MSE is calculated using the recursive window method. We first estimate the coefficients of the models using the full sample. We then let the entire period of the dataset = [1,T] = [1,R] + [R+1, T]. We use the data from [1,R] to predict R+1. We repeat the process by expanding the input data to [1, R+1] to predict R+2. (Note that R+1 is the recently predicted data.) We continue this process until the end of the data is reached. The MSE is then calculated using the out-of-sample prediction data compared to the actual data. The initial window (in-sample) contains the data up to December 2011 (i.e., R = December 2011).[14]

The regression results for the labor market sector indicators are shown in Tables 3A and 3B. Columns 1 to 4 of Table 3A display the results for unemployment registration (dismissed workers) under the four specifications, respectively. The lag variable is positive and significant under all specifications. The Google Trends variable is positive and significant under the model without the explanatory variables (Column 3). The significance of Google Trends remains robust when other explanatory variables are included (Column 4). The MPI variable is negative and significant, indicating a negative relationship between manufacturing activities and unemployment registration (dismissed workers). The full-specification model with Google Trends (Column 4) has the highest Adjusted R-Squared. In addition, this model has the lowest AIC, BIC, MSE, and out-of-sample MSE. Thus, in this case, including the Google Trends variable improves the nowcasting model in all measures of prediction accuracy.

---

[12] Note that there are some inconsistencies between the earlier data series (1987-2010) maintained by the Bank of Thailand and the later data series (2011-2017) maintained by the Ministry of Industry. First, the base years are different. The earlier data series used 2000 as the base year, whereas the latter used 2010. Second, the data during August-December 2010 is missing. Therefore, we imputed the August-December 2010 data from information obtained from the Ministry's monthly reports (which usually indicate how the index has changed from the same month of the previous year). In addition, we converted the earlier data series to match the base year used in the later data series using information obtained from the Ministry's monthly report and the existing data that was available. We acknowledge that this imputation of the data may not be perfect. However, since the official conversion method of the data is not available, this methodology appears to be the best option available given that we only want to know how manufacturing activities changed over the specified time period.

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

[14] See Carriere-Swallow and Labbe (2013) for a full explanation of the method.

Table 3A: Regression Results (Labor Market Sector)

VARIABLES	(1) LnSSLaidOffD12	(2) LnSSLaidOffD12	(3) LnSSLaidOffD12	(4) LnSSLaidOffD12
L.LnSSLaidOffD12	0.8244***	0.7248***	0.7529***	0.6909***
	(0.0625)	(0.0632)	(0.0593)	(0.0621)
LnGG_DismissedD12			0.1554*	0.1362*
			(0.0820)	(0.0704)
LnCPID12		-1.6931		-1.5685
		(1.2495)		(1.1422)
LnINTD12		0.0589		0.1106
		(0.0819)z		(0.0729)
LnAgri_IndexD12		0.5159		0.4697
		(0.3243)		(0.3322)
LnMPID12		-0.8644***		-0.7427***
		(0.2503)		(0.2398)
Constant	0.0251	0.0947**	0.0170	0.0824**
	(0.0289)	(0.0371)	(0.0298)	(0.0368)
Observations	124	124	124	124
R-squared	0.6897	0.7453	0.7081	0.7573
Model	No GG	No GG	GG	GG
Period	1/2005-5/2016	1/2005-5/2016	1/2005-5/2016	1/2005-5/2016
Adj R-Squared	0.687	0.734	0.703	0.745
MSE	0.0972	0.0825	0.0922	0.0792
Out-of-sample MSE	0.3267	0.1582	0.2141	0.1477
AIC	64.80	48.30	59.20	44.30
BIC	70.40	65.30	67.70	64.10

Note:Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Since the earliest data available for the agricultural production index is January 2005,
the time period for the empirical analysis conducted under this section started from January 2005.
Source: Authors' estimations

Table 3B displays the results for the unemployment rate. For this variable, we estimate the models using the actual data (not differenced) shown in Columns 1 to 4 and using the differenced data shown in Columns 5 to 8. Without differencing, the Google Trends variable is positive and significant. CPI, policy interest rate, and agricultural production indices have a negative relationship with unemployment rate. The full-specification model with Google Trends (Column 4) has the highest Adjusted R-Squared and lowest MSE and AIC. However, since the Augmented Dicky Fuller test revealed that the data may be affected by both seasonal factors and non-stationary issues, the non-differenced model may not be appropriate. Columns 5 to 8 display the results of the differenced models (per equations 1 to 4). Interestingly, the Google Trends variable is negative and slightly significant under the full-specification differenced model with Google Trends (Column 8). This could be due to the fact that the unemployment data series may not fit well with the model.

Table 3B: Regression Results (Labor Market Sector) Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	LnUnemp	LnUnemp	LnUnemp	LnUnemp	LnUnempD12	LnUnempD12	LnUnempD12	LnUnempD12
L.LnUnemp	0.7850***	0.5006***	0.6566***	0.4864***				
	(0.0471)	(0.0754)	(0.0626)	(0.0754)				
LnGG_Apply			0.4023***	0.2216*				
			(0.1125)	(0.1185)				
LnCPI		-1.2758***		-0.9468**				
		(0.4044)		(0.4112)				
LnINT		-0.1158**		-0.0905*				
		(0.0455)		(0.0477)				
LnAgri_Index		-0.1520**		-0.1009				
		(0.0627)		(0.0658)				
LnMPI		-0.2957		-0.3170				
		(0.2563)		(0.2517)				
L.LnUnempD12					0.6483***	0.5527***	0.6522***	0.5274***
					(0.0770)	(0.0902)	(0.0773)	(0.0886)
LnGG_ApplyD12							-0.0581	-0.2861*
							(0.1192)	(0.1484)
LnCPID12						-1.0610		-0.8466
						(0.9804)		(0.9460)
LnINTD12						-0.0676		-0.1391**
						(0.0549)		(0.0646)
LnAgri_IndexD12						0.1450		0.1228
						(0.1876)		(0.1865)
LnMPID12						-0.1823		-0.2232
						(0.2176)		(0.2046)
Constant	0.0009	7.9615***	-1.5551***	5.4499***	-0.0123	0.0045	-0.0149	-0.0171
	(0.0192)	(1.6185)	(0.4332)	(1.9657)	(0.0187)	(0.0276)	(0.0198)	(0.0304)
Observations	148	148	148	148	136	136	136	136
R-squared	0.6527	0.7035	0.6845	0.7104	0.4247	0.4505	0.4256	0.4641
Model	No GG	No GG	GG	GG	No GG	No GG	GG	GG
Period	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017	1/2005-5/2017
Adj R-Squared	0.650	0.693	0.680	0.698	0.420	0.429	0.417	0.439
MSE	0.0514	0.0451	0.0470	0.0444	0.0489	0.0482	0.0492	0.0473
Out-of-sample MSE	0.1298	0.2639	0.1387	0.2579	0.1317	0.1579	0.1577	0.1641
AIC	-17.40	-32.80	-29.60	-34.30	-22.50	-20.70	-20.70	-22.10
BIC	-11.40	-14.80	-20.60	-13.30	-16.60	-3.200	-11.90	-1.700

Note: Robust standard errors in parentheses ,*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Since the earliest data available for the agricultural production index is January 2005,
the time period for the empirical analysis conducted under this section started from January2005.
Source: Authors' estimations

3.2 The Real Sector

The real sector of the economy is associated with the production of goods and services. Within the real sector, automobile production and sales activities are the main areas focused on in this study. The automotive sector is one of the most important sectors in Thailand. As of 2016, Thailand is ranked as the twelfth largest producer of motor vehicles (passenger cars and commercial vehicles), with an overall production of 1,944,417 vehicles.[15] Similar to the labor market sector, some activities within the real sector, such as automobile sales, appeared to have shifted towards online platforms. Many official car dealers have established websites so that customers can check 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 reported having more than 85,000 automobiles currently listed.[16]

Table 4 provides summary statistics of the real sector variables used in this study. Specifically, the car brands of interest are Honda, Mitsubishi, Mazda and Toyota. The minimum, maximum and average sales volumes of these car brands during our study period are shown in the table. 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 products they intend to buy. Although, our hypothesis here is that Google Trends is linked to actual new automobile sales, we acknowledge that it is possible for Google Trends to pick up public interest from automobile discussion forums or second-hand automobile market activities.[17] We calculate the correlations of the monthly sales volume of each car brand and 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 was 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 for which the Google Trends data is available we started our data series from then through until May 2017.[21] Table 5 displays the correlations generated. It appears that brands in the Thai language have higher correlations (with actual sales volume) than brands in the English language. Therefore, we will use the brands in Thai, namely, "ฮอนด้า" (Honda), "มิตซูบิชิ" (Mitsubishi), "มาสด้า" (Mazda) and "โตโยต้า" (Toyota), as the keywords for our empirical analyses. We show trends from Google versus the trends of actual monthly sales volume of these car brands in Figures 4, 5, 6, and 7.

---

[15] The International Organization of Motor Vehicle Manufacturing (2016), retrieved from http://www.oica.net/category/production-statistics/ (as of 3 August 2017).
[16] rod.kaidee.com (as of 22 August 2017)
[17] For Thailand, the official statistics are only available for new car sales. There are no official statistics for used car sales. Therefore, in our empirical analysis we only examined the market for new car sales.
[18] The term automobile here comprises 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 finance, banking, production, investment, etc.
[20] Toyota Motor Thailand, Co. Ltd compiles and updates new automobile sales volumes (i.e., number of vehicles sold) for all leading brands in Thailand.
[21] Google Trends data was accessed during July-August 2017.

Table 4: Summary Statistics of the Variables (Real Sector: Automobile Sales)

Variable Name	Variable Description	Duration	Min	Max	Average
Sales (Honda)	Monthly Sales Volume for Honda	Jan 2004 - May 2017	332	28,708	8,695
Sales (Mitsubishi)	Monthly Sales Volume for Mitsubishi	Jan 2004 - May 2017	905	14,836	4,539
Sales (Mazda)	Monthly Sales Volume for Mazda	Jan 2004 - May 2017	632	7,702	2,659
Sales (Toyota)	Monthly Sales Volume for Toyota	Jan 2004 - May 2017	4,016	48,979	25,338

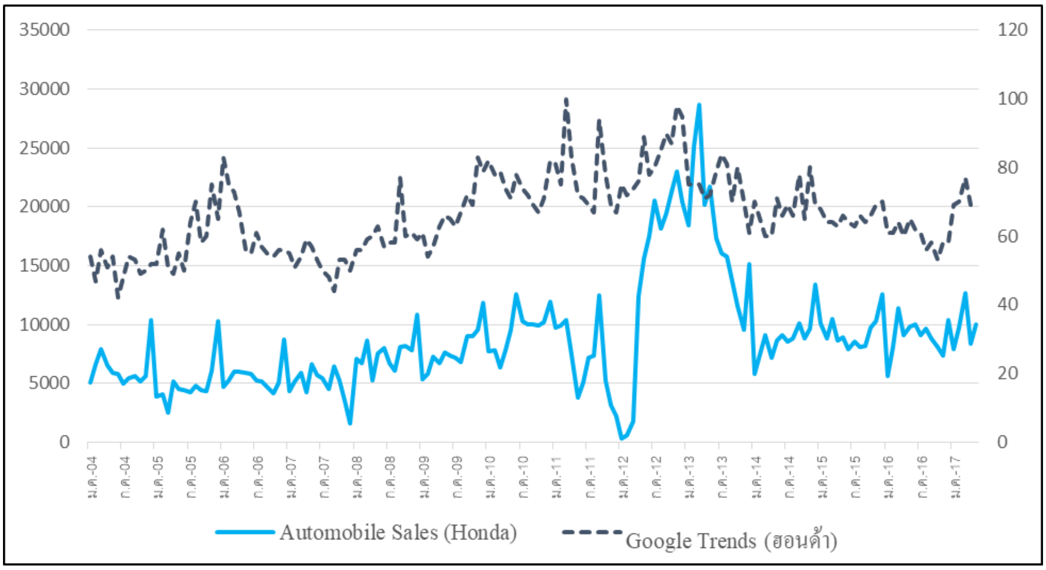
Source: Authors' calculations

Table 5: Potential Keywords (Real Sector: Automobile Sales)

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

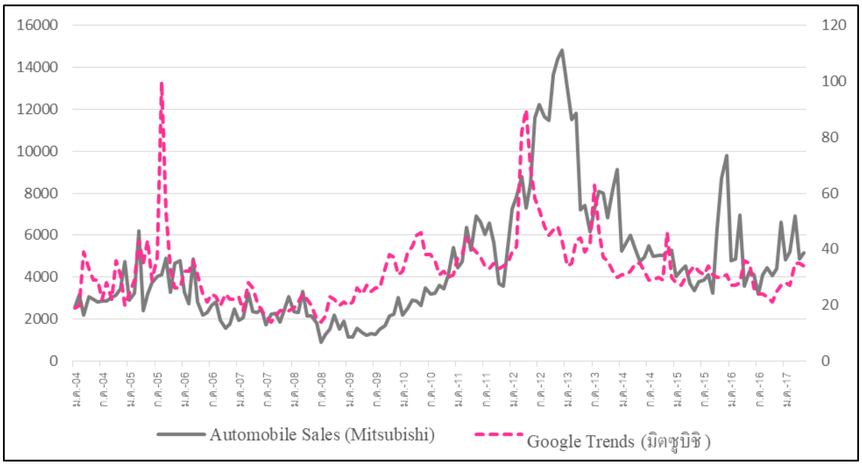
*English translation in parentheses (if applicable)
Source: Authors' calculations

Figure 4: Automobile Sales (Honda) and Google Trends



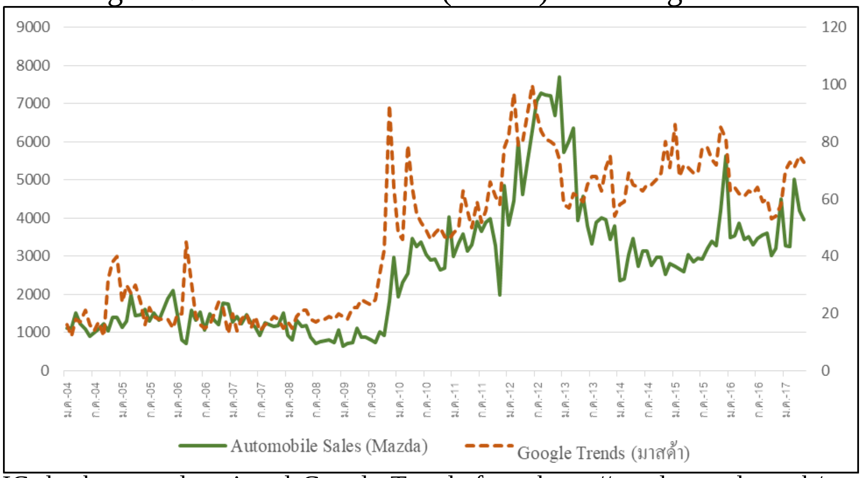
Source: CEIC database, and retrieved Google Trends from https:// trends.google.co.th/trends/ ?geo=TH with the keywords.

Figure 5: Automobile Sales (Mitsubishi) and Google Trends



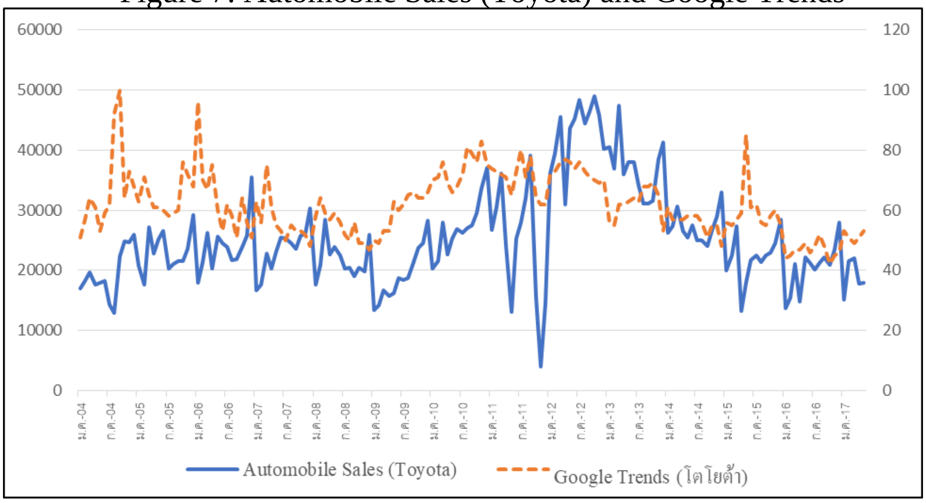
Source: CEIC database, and retrieved Google Trends from https://trends.google.co.th/trends/ ?geo=TH with the keywords.

Figure 6: Automobile Sales (Mazda) and Google Trends



Source: CEIC database, and retrieved Google Trends from https://trends.google.co.th/trends/ ?geo=TH with the keywords.

Figure 7: Automobile Sales (Toyota) and Google Trends



Source: CEIC database, and retrieved Google Trends from https://trends.google.co.th/trends/ ?geo=TH with the keywords.

The empirical analyses for this section follow the main equations discussed in the Labor Market section. Specifically, Equations 1 to 4 will be estimated, but with different variables of interest. Under this section, y_t is the natural log of the monthly sales volume for (i) Honda, (ii) Mitsubishi, (iii) Mazda, or (iv) Toyota. Similar to the previous section, the term $\Delta_{12}y_t$ (year-on-year change of the variable) is used in the model to mitigate for any seasonality effects and the non-stationary issues that may occur within the data series.

The x_t variable which is the vector of the explanatory variables, represents (i) the natural log of the Consumer Price Index (CPI), (ii) the natural log of the policy interest rate, and (iii) the natural log of the Manufacturing Production Index (MPI). These variables are selected due to the fact that price levels generally affect sales. In addition, overall manufacturing activities should be related to automobile manufacturing activities.

We obtained the CPI data from the Bureau of Trade and Economic Indices, Ministry of Commerce and policy interest rate data from the Bank of Thailand, together with MPI data from the Bank of Thailand and the Office of Industrial Economics, Ministry of Industry.[22]

For this sector, G_t is the natural log of monthly Google Trends for (i) "ฮอนด้า" (Honda), (ii) "มิตซูบิชิ" (Mitsubishi), (iii) "มาสด้า" (Mazda) and (iv) "โตโยต้า" (Toyota). $\Delta_{12}\mathrm{G}_t$ is $\mathrm{G}_t - \mathrm{G}_{t-12}$ or the year-on-year change of the variable. The time period for the analysis is from January 2004 to May 2017.[23] Robust standard errors are used in all models.

To compare the forecast accuracy among models, we examine different types of prediction errors namely, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Squared Error (MSE) and Out-of-sample Mean Squared Error (Out-of-sample MSE).[24] (See detailed explanation of the Out-of-sample MSE estimation in Section 3.1).

The regression results are shown in Tables 6A and 6B. Columns 1-4 of Table 6A display the results for Honda. The lag variable is positive and significant under all specifications. The MPI variable is positively associated with sales volume. Thus, Honda sales volume is high when manufacturing productivity is high. The Google Trends variable is positive and significant (Column 3). The significance of the Google Trends variable remains robust once the explanatory variables are included (Column 4). The full-specification model with Google Trends (Column 4) has the highest Adjusted R-Squared. In addition, this model has the lowest AIC, MSE, and Out-of-sample MSE. Thus, in this case, including the Google Trends variable improves the nowcasting model in most measures of prediction accuracy.

Columns 5-8 of Table 6A display the results for Mitsubishi and Columns 1-4 of Table 6B those for Mazda. The lag variable is positive and significant under all specifications. Similar to the Honda case, the MPI variable is positively associated with sales volume. In addition, the CPI variable is negatively associated with sales volume in the full-specification models. The Google Trends variables are positive and significant under all models (Columns 7,8 of Table 6A and Columns 3,4 of Table 6B). The full-

---

[22] Note that there are some inconsistencies between the earlier data series (1987-2010) maintained by the Bank of Thailand and the later data series (2011-2017) maintained by the Ministry of Industry. See Footnote 12 for details on how we imputed the missing data and converted the remaining data.

[23] Google Trends data was accessed during July-August 2017.

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

specification models with Google Trends (Column 8 of Table 6A and Column 4 of Table 6B) have the highest Adjusted R-Squared and the lowest AIC, MSE, and Out-of-sample MSE. Thus, these full-specification models with Google Trends perform better than other models according to most measures of prediction accuracy.

Columns 5-8 of Table 6B display the results for Toyota. The lag variable is positive and significant under all specifications. The CPI variable is negatively associated with sales volume, whereas the policy interest rate and the MPI are positively associated with sales volume. For Toyota, although the Google Trends variable is positive and significant under the model without the explanatory variables (Column 7), it is no longer significant when explanatory variables are included (Column 8). Therefore, in the case of Toyota Google Trends may just reflect information that is already provided by other explanatory variables.

Table 6A: Regression Results (Real Sector: Automobile Sales)

VARIABLES	(1) Honda LnSalesD12	(2) Honda LnSalesD12	(3) Honda LnSalesD12	(4) Honda LnSalesD12	(5) Mitsubishi LnSalesD12	(6) Mitsubishi LnSalesD12	(7) Mitsubishi LnSalesD12	(8) Mitsubishi LnSalesD12
L.LnSalesD12	0.7878***	0.7409***	0.7538***	0.7153***	0.8330***	0.7834***	0.7780***	0.7132***
	(0.0974)	(0.0960)	(0.1039)	(0.1025)	(0.0474)	(0.0493)	(0.0497)	(0.0519)
LnGG_HondaD12			0.5372**	0.4414*				
			(0.2094)	(0.2360)				
LnGG_MitsubishiD12							0.2321***	0.2451***
							(0.0618)	(0.0615)
LnCPID12		-0.8048		-1.7682		-1.9147		-2.1329*
		(1.7130)		(1.8279)		(1.2651)		(1.1745)
LnINTD12		-0.0487		-0.0464		0.0405		0.0860
		(0.1012)		(0.1002)		(0.0640)		(0.0593)
LnMPID12		1.4595***		1.3983***		0.6527***		0.6862***
		(0.4023)		(0.4029)		(0.1862)		(0.1754)
Constant	0.0090	-0.0185	0.0004	-0.0010	0.0062	0.0314	0.0082	0.0379
	(0.0431)	(0.0488)	(0.0432)	(0.0486)	(0.0226)	(0.0403)	(0.0217)	(0.0382)
Observations	148	148	148	148	148	148	148	148
R-squared	0.6211	0.6643	0.6327	0.6712	0.6936	0.7147	0.7181	0.7408
Model	No GG	No GG	GG	GG	No GG	No GG	GG	GG
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
Adj R-Squared	0.619	0.655	0.628	0.660	0.691	0.707	0.714	0.732
MSE	0.278	0.252	0.272	0.248	0.0780	0.0742	0.0723	0.0678
Out-of-sample MSE	1.2355	1.0657	1.1330	0.9979	0.2570	0.2932	0.1313	0.1260
AIC	232.7	220.8	230.1	219.7	44.50	39.90	34.10	27.70
BIC	238.7	235.8	239.1	237.7	50.50	54.90	43.10	45.70

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Source: Authors' estimations

Table 6B: Regression Results (Real Sector: Automobile Sales) Continued

VARIABLES	(1) Mazda LnSalesD12	(2) Mazda LnSalesD12	(3) Mazda LnSalesD12	(4) Mazda LnSalesD12	(5) Toyota LnSalesD12	(6) Toyota LnSalesD12	(7) Toyota LnSalesD12	(8) Toyota LnSalesD12
L.LnSalesD12	0.8446***	0.7988***	0.7830***	0.7094***	0.6524***	0.3573***	0.6237***	0.3558***
	(0.0552)	(0.0660)	(0.0558)	(0.0733)	(0.1240)	(0.0786)	(0.1193)	(0.0785)
LnGG_MazdaD12			0.2484***	0.2751***				
			(0.0641)	(0.0767)				
LnGG_ToyotaD12							0.4009***	0.1217
							(0.1373)	(0.1044)
LnCPID12		-1.1636		-2.2541**		-2.5894**		-2.5736**
		(1.2556)		(1.0596)		(1.0549)		(1.0754)
LnINTD12		-0.0022		0.0941		0.0905*		0.0944**
		(0.0788)		(0.0727)		(0.0468)		(0.0473)
LnMPID12		0.4515*		0.5496***		1.9861***		1.9360***
		(0.2302)		(0.2028)		(0.3358)		(0.3452)
Constant	0.0148	0.0315	-0.0029	0.0361	-0.0000	-0.0022	0.0106	0.0022
	(0.0225)	(0.0300)	(0.0220)	(0.0265)	(0.0256)	(0.0294)	(0.0247)	(0.0301)
Observations	148	148	148	148	148	148	148	148
R-squared	0.7134	0.7243	0.7505	0.7652	0.4254	0.6748	0.4548	0.6773
Model	No GG	No GG	GG	GG	No GG	No GG	GG	GG
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
Adj R-Squared	0.711	0.717	0.747	0.757	0.422	0.666	0.447	0.666
MSE	0.0678	0.0666	0.0594	0.0571	0.0980	0.0566	0.0936	0.0566
Out-of-sample MSE	0.2097	0.2266	0.1065	0.0863	0.2527	0.1354	0.2352	0.1344
AIC	23.70	23.90	5.200	2.200	78.20	0	72.40	0.800
BIC	29.70	38.90	14.20	20.20	84.20	15	81.40	18.80

Note: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1
Source: Authors' estimations

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.[25]

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

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

Market Value represents the current market value of all stocks whereas Base Market Value represents the market value of all stocks on 30 April 1975 (when SET was established). Table 7 provides summary statistics of the SET index during our study period.

Table 7: Summary Statistics of the Variables (Financial Sector)

Variable Name	Variable Description	Duration	Min	Max	Average
SET	Monthly SET Index	Jan 2004 - May 2017	402	1,598	1,020

Source: Authors' calculations

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

We identified potential keywords that may be entered by people interested in stock trading and looking for information about the SET index. These keywords (and their corresponding English translation, when applicable) are shown in Table 8. The table also shows correlations of these keywords with the SET index data.[26] Since January 2004 is earliest month for which Google Trends data is available, we ran our data series from then until May 2017.[27] Among the potential keywords, the keyword "หุ้น" (stock) has the highest correlation (0.9027) with the SET index. Therefore, we used this keyword for our empirical analyses. We contrasted the monthly SET index data with Google Trends of the keyword "หุ้น" (stock) in Figure 8.

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

[26] The SET index data was retrieved from https://www.set.or.th/en/market/market_statistics.html (as of 11 July 2017).

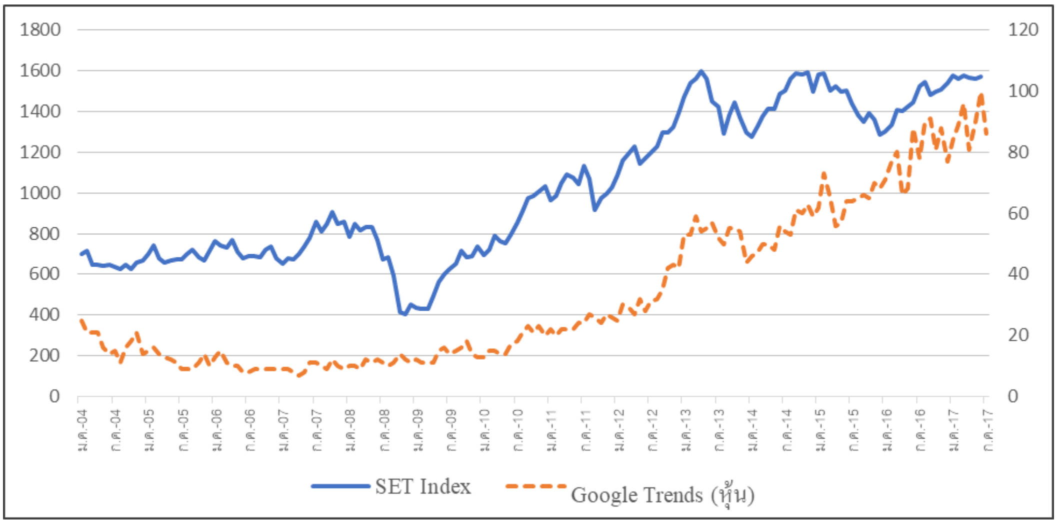
[27] Google Trends data was accessed during July-August 2017.

Table 8: Potential Keywords (Financial Sector)

Potential Keywords*	Correlations SET Index
หุ้น (Stock)	0.9027
ราคาหุ้น (Stock Price)	0.8988
ตลาดหุ้น (Stock Market)	0.7490
SET	-0.0021
SET Index	0.6779

Note: *English translation in parentheses (if applicable)
Source: Retrieved from https:// trends.google.co.th /trends/?geo=TH with the keywords.

Figure 8: SET Index and Google Trends



Source: The SET index data was retrieved from https:// www.set.or.th/en/market/ market_ statistics.html (as of 11 July 2017)

Similar to the data series for the labor market and real sectors, the data series for SET index also has non-stationary issues. On the other hand, seasonality effects do not seem to be prominent as in the previous cases. Therefore, the selected base model for the monthly SET index is in the following form:[28]

$$\Delta y_t = a + b_1 \Delta y_{t-1} + \varepsilon_{it} \qquad (5)$$

y_t is the variable of interest, which is the natural log of the monthly SET index data. t is the time variable which is month. Δy_t is $y_t - y_{t-1}$; and Δy_{t-1} is $y_{t-1} - y_{t-2}$. ε_{it} is the error term.

Since there could be other external factors that affect the SET index, we also estimate the base model with additional explanatory variables as follows:

[28] We tried many variations of AR models and attempted to select the best-fitting alternative. However, for the SET index it turned out that once we take the difference on the data, the lag term no longer explains the data. Nevertheless, we showed that including the Google Trends variable can improve the model.

$$\Delta y_t = a + b_1 \Delta y_{t-1} + b_2 \Delta x_t + \varepsilon_{it} \quad (6)$$

x_t is the vector of additional explanatory variables, namely, (i) the natural log of the Consumer Price Index (CPI), (ii) the natural log of the policy interest rate, (iii) the natural log of the market price-to-earnings (P/E) ratio, (iv) the natural log of the market dividend yield, (v) the natural log of the VIX index, and (vi) the natural log of Five-year Thai Sovereign CDS spread. These variables are standard predictors for stock return models. (See Fama (1981) for further discussion.) $\Delta \mathrm{x}_t$ is $\mathrm{x}_t - \mathrm{x}_{t-1}$ or the one period change of each of the variables.

We accessed the CPI data from the Bureau of Trade and Economic Indices, Ministry of Commerce and the policy interest rate data from the Bank of Thailand. The market price-to-earnings (P/E) ratio data and the market dividend yield data were obtained from the Stock Exchange of Thailand's website. VIX index data and the Five-year Thai Sovereign CDS spread data were both obtained from Bloomberg.

The model with Google Trends is as follow:

$$\Delta y_t = a + b_3 \Delta G_t + \varepsilon_{it} \quad (7)$$
$$\Delta y_t = a + b_2 \Delta x_t + b_3 \Delta G_t + \varepsilon_{it} \quad (8)$$

G_t is the Google Trends for "หุ้น" (Stock). $\Delta \mathrm{G}_t$ is $\mathrm{G}_t - \mathrm{G}_{t-1}$. The time period for the analysis is from January 2004 to May 2017.[29] Robust standard errors are used in all models.[30]

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), and the Out-of-sample Mean Squared Error (Out-of-sample MSE).[31] (See detailed explanation of the Out-of-sample MSE estimation in Section 3.1).

The regression results are shown in Table 9. Columns 1 and 2 display the results for the SET index without the Google Trends variable. The lag difference variable is not significant under both specifications. The dividend yield variable and the VIX index variable are negatively associated with the SET index (Column 2). These variables remain negative and significant once the Google Trend variable is used in place of the lag difference variable (Column 4). Under this full-specification model with Google Trends, the Google Trends variable is positive and significant. The model also has the lowest AIC, BIC, and Out-of-sample MSE. Thus, the model provides better prediction accuracy compared to other models according to various measures.

[29] Google Trends data was accessed during July-August 2017.
[30] Note that the full-specification ARCH/GARCH models are not robust.
[31] These are the measures commonly used in the literature. See Choi and Varian (2009a, 2009b, 2012) and McLaren and Shanbhogue (2011), for examples.

Table 9: Regression Results (Financial Sector)

VARIABLES	(1) LnSETD1	(2) LnSETD1	(3) LnSETD1	(4) LnSETD1
L.LnSETD1	0.1828	0.0432		
	(0.1151)	(0.0392)		
GG_StockD1			0.0010	0.0006*
			(0.0006)	(0.0003)
LnCPID1		0.5165		0.7347
		(0.4550)		(0.4710)
LnINTD1		0.0112		0.0101
		(0.0335)		(0.0335)
LnPED1		0.0749		0.0821
		(0.0589)		(0.0588)
LnDYD1		-0.6275***		-0.6184***
		(0.1107)		(0.1113)
LnVIXD1		-0.0302**		-0.0314***
		(0.0118)		(0.0117)
LnCDSD1		-0.0287		-0.0279
		(0.0187)		(0.0194)
Constant	0.0040	0.0049**	0.0046	0.0048**
	(0.0048)	(0.0022)	(0.0045)	(0.0022)
Observations	159	159	160	160
R-squared	0.0334	0.7845	0.0061	0.7802
Model	No GG	No GG	GG	GG
Period	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017	1/2004-5/2017
Adj R-Squared	0.0273	0.775	-0.000210	0.770
MSE	0.00328	0.00076	0.00335	0.00077
Out-of-sample MSE	0.00139	0.00039	0.00136	0.00038
AIC	-456.5	-683.1	-455.8	-685.2
BIC	-450.4	-658.6	-449.7	-660.6

Note: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1
Source: Authors' estimations

4. Conclusions

This paper illustrated how Google Trends can be used to improve predictions concerning various Thai economic indicators. Specifically, the paper utilized the real-time aspect of Google Trends to conduct nowcasting analyses – using current period real-time information to estimate current period economic indicators. The authors performed nowcasting analyses in three areas, 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 both the Adjusted R-Squared and prediction accuracies in terms of various measures. Our results are in line with Choi and Varian (2009a, 2009b, 2012) used Google Trends to nowcast similar economic indicators in the United States. Our results are also in line with other studies conducting analyses in advanced countries, such as the United Kingdom (McLaren and Shanbhogue (2011)), Germany (Askitas and Zimmermann (2009)) and France (Fonduer and Karame (2013)) and in emerging middle income countries, including Chile (Carriere-Swallow and Labbe (2013), Turkey

(Chadwick and Sengul (2012) and Zeybek and Ugurlu (2015)) and Central American nations (Seabold and Coppola (2015)).

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

However, what this study tries to argue is that, despite these drawbacks, information retrieved from Google searches can still be shown to be notably useful in many cases. With Thailand, Google Trends data was invaluable in nowcasting various economic indicators in three spheres of interest, (i) the labor market sector, (ii) the real sector, and (iii) the financial sector. As already mentioned, Google is currently the most broadly 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 is too important to be ignored.

In the future, economics research will be driven more and more by data. In the age of the digital economy, the new major source of data for research is data from the internet, including Google Trends and many other sources. 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 available data to help modify existing methodologies currently being applied. Perhaps, shortcomings in the data can be fixed and the efficiency of how such data is processed can be improved. Under this environment, many more meaningful research questions can be asked and many more rigorous analyses in myriad contexts can be conducted.

Acknowledgement

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. In addition, the authors are grateful for the comments from the two anonymous referees and the participants at the 4^{th} Trilateral Workshop in Economics: Applied Economics (organized by CU-TU-NTU) in 2017. The results and findings reflect the personal views of the authors and do not reflect the views of their professional institutions.

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. *Working Paper* No. 12/18, Central Bank of the Republic of Turkey.

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(s1), 2-9.

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *The American economic review*, 71(4), 545-565.

Fonduer, Y., & Karame, F. (2013). Can Google data help predict French youth unemployment?. *Economic Modelling*, 30(1), 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, Puey Ungphakorn Institute for Economic Research.

McLaren, N. & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, 2011Q2, 134-140, Bank of England.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, 25(100), 283-299.

Seabold, S., & Coppola, A. (2015). Nowcasting prices using Google trends: an application to Central America. *World Bank Group Policy Research Working Paper* No.7398, World Bank.

Suhoy, T. (2009). Query indices and a 2008 downturn: Israeli data. *Bank of Israel Discussion Paper* No. 2009.06, Bank of Israel.

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(1), 132-139.

Zeybek, O. & Ugurlu, E. (2015). Nowcasting credit demand in turkey with Google trends data. *Journal of Applied Economic Science*, 10(2), 293-300.